Cleavage of Potato Virus Y Polyprotein in *Escherichia coli* Depending on the Proteolytic Activity of Viral Protease

Makoto Hidaka, Yukiko Yoshida, Haruhiko Masaki, and Takeshi Uozumi

Department of Biotechnology, Faculty of Agriculture, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan
Received October 8, 1992

We have cloned a 5-kb cDNA corresponding to the 3'-half of genomic RNA of potato virus Y (PVY), the type member of plant potyvirus. Based on its nucleotide sequence, the cDNA was presumed to encode PVY protease responsible for the self-processing of polyprotein precursor expressed from PVY genome. To test its proteolytic activity, the cDNA was modified so as to express the protease-polymerase-coat protein (CP) portion of PVY polyprotein in *Escherichia coli*. By Western blotting analysis of the cell extract of *E. coli* expressing the polyprotein, mature CP was detected. A large deletion in the polymerase-coding region did not affect the emergence of mature CP, but a small deletion in the putative protease region resulted in disappearance of mature CP and appearance of the intact polyprotein. These results indicated that the polyprotein expressed in *E. coli* was cleaved depending on the proteolytic activity of PVY protease.

Potato virus Y (PVY) is the type member of the potyvirus group, which contains more than 60 distinct members and is the largest plant virus group. The potyviral particles consist of up to 2000 subunits of a single coat protein and a single molecule of single-stranded RNA. The genomic RNA is approximately 10 kb long and is supposed to be translated into a large polyprotein, which is then processed to a coat protein and at least eight viral non-structural proteins by its own protease activity. Studies on the self-processing of the polyprotein of tobacco etch virus (TEV), a member of the potyvirus group, demonstrated that the small nuclear inclusion protein (49K protein or N1a protein) portion of the polyprotein has the proteolytic activity that is responsible for several cleavages of the polyprotein. This viral protease plays the main role during the self-processing of the TEV polyprotein. Similar proteolytic activities were shown in the N1a proteins of two other potyviruses, tobacco vein mottle virus and plum pox virus. On the other hand, though a full-length cDNA clone and partial cDNA clones of several PVY strains have been isolated and their nucleotide sequences were already analyzed, the proteolytic activity of the PVY polyprotein has not yet been reported.

We cloned a 5.0-kb cDNA of PVY genomic RNA and sequenced its nucleotides. Based on this sequence, we presumed that the cDNA encoded the C-terminal half of PVY polyprotein, which should be cleaved into five viral proteins including the coat protein. The viral protein of 50 kDa, which was tentatively designated "protease", encoded in this cDNA was expected to be responsible for this self-processing of the PVY polyprotein. To demonstrate its proteolytic activity, we introduced plasmids containing PVY cDNA fragments with various deletions into *Escherichia coli* cells and analyzed the structures of the expressed PVY polyproteins by Western blotting analyses using anti-PVY coat protein antibody and anti-PVY "protease" antibody. In this paper, we show that this "protease" cleaves the PVY polyprotein site-specifically. And we also indicate that the N-terminal portion of the "protease" is not necessary for the proteolytic activity.

Materials and Methods

*Bacterial strain. Escherichia coli* GC4670 (lon::Tn10 thr leu lacY) was used as the host cell for the expression of PVY polyproteins. This strain was obtained from the *E. coli* collection of National Institute of Genetics (Japan) under the strain number ME8417.

*PVY cDNA. Cloning and sequencing of a 5.0-kb cDNA of PVY genomic RNA were described in a previous paper. The nucleotide sequence is in the DDBJ, EMBL, and GenBank data base under the accession number D12539. The 4667-bp region of the cDNA encodes the C-terminal half of PVY protease. We presumed that the open reading frame (ORF) for this polyprotein consisted of five portions: 1 to 857 nucleotide (nt) portion encoding the C-terminal half of the cytoplasmic inclusion body, 858 to 1013 nt encoding 5K protein, 1014 to 2309 nt encoding protease (50-kDa protein), 2310 to 3866 nt encoding polymerase, and 3867 to 4667 nt encoding coat protein (see Fig. 1A).

*Antibodies. Anti-PVY coat protein antibody was purchased from Boehringer Mannheim. Anti-PVY 50-kDa protein antibody was prepared as follows. An 1128-bp Kpnl–CiaI fragment (1155 to 2283 nt) of the PVY cDNA was blunt-ended with the Klenow fragment of *E. coli* DNA polymerase I, ligated to BamHI linkers, and inserted into the BamHI site at the 3'-end of lacZ gene of pUEX1. *E. coli* HB101 harboring this plasmid was incubated at 42°C for 1 h then at 37°C for 1 h. The cells were harvested and disrupted by sonication. Inclusion bodies formed in the cells were collected by centrifugation at 2500 × g for 10 min and dissolved with 50 mM Tris–HCl (pH 7.5) containing 8 M urea. The fractions containing "β-galactosidase-50-kDa protein" fusion protein of 150 kDa were collected, dialyzed against water, and then lyophilized. Rabbits were immunized with the fusion protein and antiserum was prepared. The immunoglobulin G (IgG) fraction was isolated from the serum using an Amjure Pure Kit (Amer sham). Anti-PVY 50-kDa protein antibody was purified from the IgG by passing it through a column of β-galactosidase- and *E. coli* proteins-coupled Sepharose, which was prepared as follows. The cells of *E. coli* HB101 harboring pUEX1 were harvested after incubation at 42°C for 1 h, then at 37°C for 1 h and a cell extract was prepared by sonication. One milliliter of CNBr-activated Sepharose 4B (Pharmacia) was incubated with the cell extract containing 5 mg proteins. The Sepharose was recovered and used to make the column. In this paper, we call this anti-50-kDa protein antibody "anti-PVY protease antibody" from the beginning according to the designation of viral proteins described in our previous paper.

*Construction of plasmids. An expression vector of *E. coli*, pUC18, was digested with KpnI. The fragment was blunt-ended with Klenow enzyme and self-ligated. The BamHI–AccI portion of the plasmid was replaced by a 4.0-kb *BclI–CiaI* fragment (928 to 4886 nt) of the PVY cDNA. The
obtained plasmid was named pPVY1. By this construction, an ORF encoding the PVY polyprotein was joined to a pUC18-derived ORF, which starts from the ATG codon of the lacZ gene and encodes only 11 amino acid residues (see Fig. 3A). Then, pPVY1 was partially digested with Fpol after complete digestion with Aafl and the generated fragments were self-ligated. A pPVY1 derivative, which lacked an 1146-kb fragment (2466 to 3612 nt) of the polymerase-encoding region, was isolated and designated pPVY2. A plasmid, pPVY4, was constructed as follows. A 958-bp fragment, including the Ncol–EcoRI portion described below, generated by EcoRI digestion of pPVY2 was chewed by incubation with BAL 31 nuclease at 30°C for 1 min, ligated with EcoRI linkers, and digested with Ncol. The ligates were electrophoresed and fragments of approximately 300 bp were recovered. Then, the 333-bp Ncol–EcoRI portion of pPVY2, corresponding to 150–2013 bp of the PVY cDNA, was replaced by the recovered fragments. Some of the obtained plasmids were sequenced around the EcoRI site and a plasmid that has a 42-bp deletion in the upstream region of the EcoRI site, i.e., in the promoter-encoding region, was selected. This 14-amino acid deletion consisted of a 15-amino acid elimination caused by the digestion with BAL 31 nuclease and an addition of a proline residue derived from the EcoRI linker. The final structure of the PVY cDNA fragment of each plasmid is shown in Fig. 1A.

Expression of PVY polyprotein in E. coli. To the culture of E. coli GC4670 harboring pPVY1, pPVY2, or pPVY4, isopropyl-β-D-thiogalacto side (IPTG) was added at the final concentration of 2 mm and the culture was incubated at 37°C for further 6 h. The cells were harvested, resuspended with 50 mm Tris–HCl (pH 7.5), and disrupted by sonication. The suspension was centrifuged at 8000 g for 10 min and the supernatant was recovered, which was called soluble protein fraction. The remaining precipitate was resuspended with 50 mm Tris–HCl (pH 7.5) and called insoluble protein fraction. The expressed PVY polyproteins were detected by Western blotting analysis. As a control, cell extracts of E. coli GC4670 harboring pUC18 were prepared and used for Western analysis.

![Diagram of the PVY cDNA](image)

**Fig. 1. Schematic Illustration of the Structure of the PVY cDNA.**

(A) The physical map of the PVY cDNA is illustrated at the top. The large box and thick line indicate open reading frame (ORF) and untranslated region, respectively. Vertical lines in the box represent the internal cleavage sites of the PVY polyprotein. The name of each viral protein was abbreviated as follows: CI, cytoplasmic inclusion body; SK, 5S RNA protein; Pro, protease; Pol, polymerase; CP, coat protein. The estimated molecular masses of Pro, Pol, and CP are shown under the name of each protein. The structures of the ORFs encoded by the cDNA inserts contained in pPVY1, pPVY2, pPVY4, and pPVY2-SD are illustrated with shadowed boxes. The black box at the left end of each ORF indicates the ORF derived from the lacZ gene. Deletions in the cDNA inserts are indicated by dashed lines. The position and the nucleotide sequence of the site-directed mutagenesis introduced into pPVY2 are marked over the box of pPVY2-SD. (B) Deduced amino acid sequence of PVY protease is shown.

The amino acid residues identical to that of TEV protease are marked by bold face. The histidine, aspartic acid, and cysteine corresponding to the amino acids constituting the catalytic triad of TEV protease are marked by dots. Underline represents the position of the 14-amino acid deletion in pPVY4.
Western blotting analysis. Thirty micrograms of proteins in the soluble fraction and fifteen micrograms of proteins in the insoluble fraction were size-fractionated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using 12% gel and blotted onto an Immobilon PVDF membrane (Millipore). The membrane was incubated with anti-PVY coat protein antibody or anti-PVY protease antibody and then incubated with 125I-labeled Protein A (Amerham). The positions of the proteins to which the antibodies reacted were detected by autoradiography.

Site-directed mutagenesis. The cDNA insert of pPVY2 was subcloned with pUC118. Mutagenesis of the plasmid was performed using the Mutan-K Site-directed Mutagenesis System (Takara). The nucleotide sequence of the synthetic mutagenic oligonucleotide was 5'-GGGCGTACCAACGGACATCAACATGTA-3' corresponding to 1172-1204 portion of the PVY cDNA. The mutation was confirmed by sequencing and the fragment was put back to pPVY2. The constructed plasmid was named pPVY2-SD.

Amino acid sequencing. The N-terminal amino acids of proteins were sequenced using a Model 477A Peptide Sequencer (Applied Biosystem).

Results and Discussion
Expression of PVY protease activity in E. coli

As we described in our previous paper, the cloned PVY cDNA encodes the C-terminal half of PVY polyprotein (Fig. 1A) and the amino acid sequence of the N-terminal region of coat protein (CP) was presumed to be ANDTIDAGGG. To confirm this prediction, we analyzed the N-terminal amino acid sequence of the authentic PVY CP obtained from the virus particles. The sequence found, ANDTI, showed complete agreement with the presumed sequence, suggesting that the predicted cleavage site between polymerase and CP is actually functional. The deduced amino acid sequence of the 50-kDa protein, designated as protease, is shown in Fig. 1B. This sequence shows approximately 50% homology with that of TEV protease (49K protein). Moreover, three amino acids, i.e., histidine, aspartic acid, and cysteine, constituting the catalytic triad of TEV protease are conserved in the sequence of PVY protease. Based on these findings, we expected that this putative PVY protease shows the proteolytic activity responsible for the cleavage of PVY polyprotein.

The plasmid pPVY1 encodes a part of PVY polyprotein (see Fig. 1A), the molecular mass of which was predicted to be 145kDa. If the polyprotein had proteolytic activity, 30-kDa CP should be released from it. The mobility of the CP on SDS-PAGE corresponded to 32kDa (Fig. 2B; lane M). The molecular mass of the polyprotein expressed from pPVY1 in E. coli was investigated by Western blotting analysis using anti-PVY coat protein antibody (anti-Cp Ab) and anti-PVY protease antibody (anti-Pro Ab). To observe the cross-reactivities of each antibody to E. coli proteins, the cell extract of E. coli GC4670 harboring pUC18 was used for Western analysis (Fig. 2B and C; lanes 7 and 8). Only one band of 40kDa was detected non-specifically by each antibody, suggesting that these antibodies can distinguish PVY proteins specifically.

We initially introduced pPVY1 into E. coli MV1184 to produce polyprotein, but the expressed polyprotein was degraded irregularly (data not shown). We thought that this degradation occurred owing to E. coli proteases and changed the host strain for E. coli GC4670, a mutant defective in the intracellular La protease. In the cells of IPTG-treated E. coli GC4670 harboring pPVY1, huge inclusion bodies were formed even two hours after the IPTG induction (data not shown). The inclusion bodies were separated from the soluble proteins by centrifugation and both soluble and insoluble protein fractions were analyzed.

**Fig. 2. Western Blotting Analysis of PVY Polyproteins Expressed in E. coli.**

 Autoradiograms of Western analysis with anti-PVY coat protein antibody and anti-PVY protease antibody are shown in panels B and C, respectively. Photograph of Coomasie Brilliant Blue-stained gel before blotting is also shown in panel A. The mobilities of molecular mass markers are indicated on the left side. In every panel, cell extracts of E. coli harboring pPVY1 (lanes 1 and 2), pPVY2 (lanes 3 and 4), pPVY3 (lanes 5 and 6), and pUC18 (lanes 7 and 8) were electrophoresed. On lanes 1, 3, 5, and 7, the soluble protein fraction (sol) of each cell extract was put. On lanes 2, 4, 6, and 8, the insoluble protein fraction (ins) of the cell extracts was put. On lane M of panel B, sap extracted form leaves of PVY-infected tobacco was electrophoresed. The proteins, the N-terminal amino acid sequences of which were analyzed, were marked with arrowheads in panel A.
by Western blotting. As shown in Fig. 2B (lane 2), not a 145-kDa band but two bands of 32 and 31 kDa were detected in the insoluble fraction by anti-CP Ab. The N-terminal amino acid sequences of the 32-kDa protein was analyzed. The sequence found, ANTDIDAGGG, was identical to that of authentic CP described above, implying that the 32-kDa protein is the mature CP. On the other hand, many bands were detected by anti-Pro Ab in spite of its specificity (Fig. 2C; lane 2). The molecular mass of the largest band was 95 kDa. These results indicated that the 145-kDa PVY polypeptide expressed in E. coli was cleaved at the internal cleavage site between polymerase and CP by its own self-processing activity. The cleavage between protease and polymerase may occur seldom in E. coli. The polypeptide with the intact form was detected neither in the soluble protein fraction nor in the insoluble protein fraction. This suggested that the expressed polypeptide molecules were cleaved quickly before they were aggregated into the inclusion bodies. Possibly, the cleavage occurs in E. coli immediately after the translation of the polypeptide.

**Deletion analysis of localization of protease activity on PVY polypeptide**

To identify the polypeptide portion showing the proteolytic activity, two deletion mutants of the PVY polypeptide were constructed. One of them, pPvY2, had a large deletion introduced in the polymerase region and the other, pPvY4, had a 14-amino acid deletion introduced in the C-terminal region of protease and the deletion in polymerase simultaneously (see Fig. 1A). The molecular masses of the polyproteins encoded by both plasmids were predicted to be 100 kDa and 98 kDa, respectively. By analysis of the cell

---

**Fig. 3. Site-directed Mutagenesis of pPvY2.**

(A) The nucleotide sequence of 5'-end region of the PVY polypeptide-encoding ORF on pPvY2 are shown. The encoded amino acids were also shown in italic type beneath the nucleotide sequence. The amino acids of the N-terminal region of PVY protease are marked by bold face and are numbered from the N-terminus of the protease. The determined N-terminal sequence of the PVY polypeptide encoded by pPvY2 is indicated by underline. Double underline represents the N-terminal sequence of the PVY polypeptide encoded by pPvY2-SD. The GGAGG sequence described in the text is marked by dashed line and the nucleotide sequence of the replacement is indicated over the line. In panels B and C, autoradiograms of Western analyses with anti-CP Ab and anti-Pro Ab are shown, respectively. The insoluble protein fractions of cell extracts prepared from E. coli harboring pPvY2 (lane 1), pPvY2-SD (lane 2), and pUC18 (lane 3) were electrophoresed. The position of CP is indicated by arrow on the left side and the mobilities of molecular mass markers are indicated on the right side. The 70-kDa protein, the N-terminal amino acid sequence of which, was marked with an arrowhead.
extract of *E. coli* harboring pPVY2, the 32-kDa band was also detected by anti-CP Ab (Fig. 2B; lane 4). This indicates that the proteolytic activity of the polyprotein is not affected by the large deletion introduced in polymerase. The other part of the polyprotein was detected as a 57-kDa protein, the strongest band among the several bands, by anti-Pro Ab (Fig. 2C; lane 4).

On the other hand, when the cell extracts of *E. coli* harboring pPVY4 were analyzed, the 32-kDa band was not observed and a band of 85 kDa was detected by anti-CP Ab (Fig. 2B; lane 6). An 85-kDa band, the largest of several bands, was also detected by anti-Pro Ab (Fig. 2C; lane 6). These results indicate that the 85-kDa protein is the intact polyprotein expressed from pPVY4, and show that the proteolytic activity of the polyprotein disappears when there is a 14-amino acid deletion introduced in the protease region.

From this, we concluded that the PVY 50-kDa protein designated as protease is really functional and responsible for the cleavage of the polyprotein between polymerase and CP. In the deleted 14 amino acid residues, the cysteine residue corresponding to that of TEV protease constituting the catalytic triad was included. Possibly, even in PVY protease, the cysteine is also important for the expression of proteolytic activity.

The molecular mass of the polyprotein expressed from pPVY4 (85 kDa) was 13 kDa smaller than that predicted from its amino acid sequences (98 kDa). This difference was observed in the case of the polypeptide released from the polyprotein expressed from pPVY2, i.e., 57 kDa (the band detected with anti-Pro Ab) and 70 kDa (the value predicted from the sequence). To discover the causes of these differences, the N-terminal amino acid sequences of the 85- and 57-kDa proteins were analyzed. These sequences were identical, MYGFDPEYTS, and coincided with the sequence starting from the 63rd amino acid residue of the protease (Fig. 3A). This N-terminal truncation of the polyproteins resulted in decreasing the molecular mass of approximately 12 kDa, which showed good agreement with the differences (13 kDa) described above. In the nucleotide sequence around the ATG codon of the 63rd methionine, a GGAGG sequence, which is very similar to the ribosomal binding site of *E. coli*, was present 10-bp upstream from the ATG codon. This strongly suggested that internal initiation of translation from this ATG codon occurred. Some other internal initiations might occur and this may be the main reason for the detection of many bands by anti-Pro Ab. Another possibility is that the degradation of the polyproteins occurred owing to *E. coli* proteases.

**Expression of longer polyprotein translated from the first initiation codon**

To avoid the internal initiation of translation from the 63rd ATG codon, the GGAGG sequence of pPVY2 was exchanged for GACGG by site-directed mutagenesis without any replacements of encoded amino acid residues. By Western analysis of *E. coli* harboring this plasmid, pPVY2-SD, a 32-kDa band was detected by anti-CP Ab (Fig. 3B; lane 2) and a 70-kDa band, the largest of several bands, was detected by anti-Pro Ab (Fig. 3C; lane 2). In Fig. 3B, some non-specific bands were detected because of changing the lot of the purchased anti-CP Ab. The amino acid sequence of the N-terminal region of this 70-kDa protein was TMITNN, which corresponded with the sequence starting from the second amino acid residue of the polyprotein encoded by pPVY2-SD. The first methionine might be detached in *E. coli*. From these lines of results, we concluded that the PVY polyprotein expressed from pPVY2-SD is cleaved by its self-processing activity into CP and 70-kDa polypeptide, two possible cleavage sites in which, i.e., the junction sites at the N and C termini of protease, are not digested in *E. coli*. Moreover, we also concluded that the N-terminal truncation of the polyprotein expressed from pPVY2 is due to the internal initiation of the translation, not the degradation by *E. coli* proteases, implying that the N-terminal 62-amino acid portion of PVY protease is not necessary for its proteolytic activity. This result is in accordance with the report by Carrington and Dougherty that only the C-terminal half of the TEV protease was required for its proteolytic activity. Possibly, the reaction mechanism of PVY protease may be similar to that of TEV protease. However, PVY protease could not cleave the polyprotein at junction sites other than that of the N terminus of CP in *E. coli*. Perhaps, the conformation of the PVY polypeptide expressed in *E. coli* is different from that of the polypeptide in plant cells and is not adequate to the recognition by the protease.

**References**