Cloning and Sequencing of Virus Inhibiting Gene Encoding an Antiviral Protein From the Leaves of Pokeweed (*Phytolacca Americana* L.)

Hikmet Murat SİPAHİOĞLU*  İlhan KAYA*  Mustafa USTA*  Sibel SAMSUN**

* Department of Plant Protection, Faculty of Agriculture, YüzüncüYıl University, 65080 Van - Turkey
** Department of Biology, Faculty of Science, YüzüncüYıl University, 65080 Van - Turkey
Corresponding author e-mail: hmsipahi@yyu.edu.tr

Accepted for publication October 10, 2013

ABSTRACT

Pokeweed antiviral protein (PAP) is single-chain ribosome inactivating protein (RIP) of *Phytolacca americana* L. (Pokeweed) that is characterized by its ability to depurinate ribosomes. In the present study, we cloned and sequenced complete gene of pokeweed antiviral protein type 1 (PAP-I) from the summer leaves of pokeweed collected from Trabzon (Turkey) using a pair of gene specific primers based on the known N- and C-terminal nucleotide sequences of PAP gene. A product of 942 base pair was purified and inserted into pGEM-T Easy vector (Promega), downstream of the T7 promoter, and transformed into *E. coli* strain, JM 109. The PAP-I cistron of *P. americana* contained 313 amino acid residues. The DNA sequence of 942 base pairs included an open reading frame (ORF). The nucleotide sequences of PAP-I gene contained no introns and the comparison with the PAP-I sequence with the PAP isoform, showed an identity of 80-99%. Sequence analysis of PAP-I revealed that it contains a single point mutation, changing the Leucine (L) at position 273 to Phenylalanine (F) (L273F) at the putative active site.

Key words: antiviral proteins, characterization, cloning, PAP-I gene, *Phytolacca americana*

INTRODUCTION

Pokeweed (*Phytolacca Americana* L.) antiviral protein (PAP) is a single chain ribosome inactivating protein (RIP) which inactivates both eukaryotic and prokaryotic ribosomes isolated from the leaves of pokeweed plants. The PAP removes specific adenine and guanine residues from the highly conserved alpha-sarcin/ricin (S/R) loop in the large rRNA (Endo et al. 1988; Hartley et al. 1993; Tomlinson et al. 1974; Ussery et al. 1977; Zarling et al. 1990). PAP is currently under clinical trials against cancer, because of its cytotoxicity to dividing cells (Waurzyniak et al. 1997; Bijal et al. 2002). PAP has also been shown to efficiently depurinate single stranded DNA (Nicolas et al. 1998), double-stranded DNA (Wang and Tumer, 1999) and adenine-containing polynucleotides (Barbieri et al. 1997). Ribosome inactivating proteins of pokeweed, PAP-I, PAP-II, and PAP-III, are different isoforms of PAP isolated from the spring leaves, early summer leaves and late summer leaves of the pokeweed plant (Irvin and Uckun 1992; Rajamohan et al. 1999).

Kataoka et al., (1991) stated that the most of the amino acid residues composing the N-terminal signal sequences are hydrophobic, which is a common feature of signal sequences of eukaryotes and prokaryotes. Today
the N-terminal amino acid composition is known for a number of RIPs. A high degree of identity (83%) was found between PAP-I and dodecantrin, from an African relative of pokeweed, than between PAP-I and PAP-S (55%), which are located in the same plant but indifferent tissues (Houston et al. 1983; Xu 1997).

All RIPs examined so far have antiviral activity against animal (Foa’-Tomasi et al. 1982) or plant viruses (Stevens et al. 1981). On the other hand, they have been considered that they are not natural antiviral agents since crude plant extracts with antiviral and RIP activity do not act against viruses on their own species (Stevens et al. 1981; Gendron and Kassanis 1954; Stirpe and Barbieri 1986). In the proposed study, we present the complete sequence of PAP-I, the major RIP component isolated from P. americana leaves by cloning the genomic DNA fragment encoding the mature PAP-I.

MATERIALS AND METHODS

Plant material. The plant material of Phytolacca americana L. for the proposed study were collected in July 2011 when the leaves were still green from the coastal area of Trabzon (Turkey). Herbarium specimens of P. americana plants were deposited at Yüzüncü Yıl University herbarium. Mature leaves of the plants were studied in both fresh and dried conditions. The care was taken to select healthy plants and for normal organs.

Preparation of genomic DNA and amplification by polymerase chain reaction. Genomic DNA from P. americana was extracted by using a commercial DNA purification kit (Bioline, Luckenwalde, Germany). PAP-specific DNA fragments were amplified from the genomic DNA by PCR using PAP-F primer (5’-ATGAAGTCTGTCTTGTGGT-3’) designed according to the N-terminal sequence of mature PAP-I, and PAP-R primer (5’-TCGAATCCTTCAAATAGATCACC-3’) designed according to the C-terminal sequence. PCR amplifications were carried out in a 50 μl reaction containing 2 μl of genomic DNA, 1 μl of each primer, 1 μl of the four dNTPs (0.25 mM) (Fermentas Life Sciences, Lithuania), 3 μl of MgCl2 (25 mM), 5 μl 10 x TaqDNA buffer (Promega, USA) and 0.4 μl of Hot Start Taq DNA polymerase (Promega, USA). PCR was performed in Master cycler thermal cycler (Eppendorf, Germany) with the following thermal cycling scheme: 2 min at 95 °C, 35 cycles of 45 s at 95 °C, 1 min at 55 °C, and 45 s at 72 °C followed by a final incubation of 5 min at 72°C. Amplified PCR products were electrophoresed in 1.5% agarose gel and stained with ethidium bromide. PCR fragments were excised under UV transilluminator and purified using Gel DNA Extraction Kit (Bioline, Germany).

Cloning and sequencing of the PAP-I gene. The purified PCR fragments of PAP-I gene was cloned into plasmid vector using TA Cloning Kit (Promega, USA) and transformed into competent Escherichia coli strain JM 109 by following manufacturer’s instructions. The transformed colonies were screened and confirmed for the presence of insert by colony PCR as described above by simply adding a single colony of bacteria or 2 μl of overnight culture to a PCR master mix instead of genomic DNA. Plasmid DNA was isolated by a commercial kit (Bioline). The purified plasmids were sequenced by automated DNA sequencer (Applied Biosystems) at Iontek Research and Biotechnology Company. DNA sequencing was performed on both strands.

Characterization of genomic traits, determination of sequence homology and phylogenetic analysis. The nucleotide sequences of the cloned PAP-I fragment was translated in to amino acid sequences using CLC Workbench Software. The deduced amino acid sequences were compared with protein sequences deposited in the GenBank using Vector NTI Software (Invitrogen). Pair wise comparison of PAP-I sequences with known PAP sequences was made using the same tool.

Signal peptide prediction. In order to predict signal peptide structure of the PAP-I protein, computer analysis were performed on the established full sequence. For this purpose, the signal peptide prediction package of CLC Main Workbench Version 6.5 (CLC bio, Denmark) was used.

Nucleotide accession number. Nucleotide sequence reported in this paper have been submitted to the GenBank/EMBL database and assigned the accession number JX446580 for the PAP-I gene of P. americana.
RESULTS AND DISCUSSION

Collection site of plant material. The plant material was collected at the beginning of July (2011) at the Black Sea Region (Turkey). Location of *Phytolacca americana* L. collection site in Trabzon along the coast line, from stony habitat at the coordinates of 50 m, N 40° 58'09.0" and E 39°50'23.2" (Figure 1).

![Image of summer flowers and seeds](image)

*Figure 1. Summer flowers (a) and the seeds (b) of *Phytolacca americana* L. collected from Trabzon province.*

Amplification and cloning of DNAs encoding PAP-I. The DNA fragment of about 942 bp encode the mature PAP-I was amplified from the genomic DNA by PCR using a pair of gene specific primers based on the N-terminal and C-terminal sequences of PAP-I (Figure 2). Although a unique alteration (Luc to Fen) in the deduced amino acid sequences of the present PAP-I were detected, the DNA sequencing revealed that the cloned PAP fragment corresponded to the DNA for PAP-I.

![Image of PCR detection of PAP gene](image)

*Figure 2. PCR detection of PAP gene of about 942 bp from the leaf (1) and the seed (2) tissue of *Phytolacca americana* L. M: Molecular size markers 3.000 bp.*

Gene structure characteristics and phylogenetic analysis. The complete amino acid sequence of PAP-I is shown in Figure 3. The deduced amino acid sequence of PAP-I gene was 75-99% identical to that deduced amino acid sequences of other RIPs from databases belonging to other *Phytolacca* species. At the N-terminus of genomic clone, the deduced amino acid sequence contained a putative signal sequence of 22 amino acids (Figure 3). The alignment analysis showed a strong homology in the conserved region (AIQMVSEAARFKYI) (Lin et al., 1991) (Figure 3). A multiple alignment analysis of the PAP-I amino acid sequence with those of PAP types showed that, majority of the residues are identical and are well conserved. The PAP-I gene has a translation initiation codon (ATG), downstream of the first base and terminated by the stop codon (UGA) located after the 939th nucleotide. The PAP-I gene contained a complete open reading frame and had no introns.
CLONING AND SEQUENCING OF VIRUS INHIBITING GENE ENCODING AN ANTIVIRAL PROTEIN FROM THE LEAVES OF POKEWEED (*PHYTOLACCA AMERICANA* L.)

The amino acid sequence was identical to that determined by others in databases, except for a single substitution of Leu 273 to Phe (Fig 4). In order to understand whether this divergence occurs within the gene or whether this small variation represents cloning artifacts, we sequenced the cloned PAP-I DNA from both directions. We identified that the PAP-I sequence of *P. americana*, collected from Trabzon, had a substitution of Thymine (T) with Cytosine (C) at position of 817 corresponding the substitution of LeuCUC (Leu) 273 to UUC (Phe) on the sequence (Figure 4). This differed the sequence from the same nucleotide positions of other PAP-I sequences. Bidirectional sequencing clearly confirmed the presence of nucleotide substitution on both sequences at the corresponding sites of DNA molecule.

**Figure 3.** The N-terminal domain of the PAP-I amino acid sequence of *Phytolacca americana* L. showing a signal peptide sequence of 22 amino acids and the position of the conserved AIQMVSEAARFYI amino acid sequence. Peptides are numbered according to their position in the polypeptide chain.

**Figure 4.** Alignment of the deduced amino acid sequence of PAP-I with other PAPs from different *Phytolacca* species. Single point mutation, at the position of 273, is shown on the amino acid chain. GenBank accession numbers of PAPs used for analysis are as follows: PAP-I (JX446580); PAP-I fusion protein synthetic construct (EF508129); PAP from *Phytolacca americana* L. (X55383); PAP from *P. americana* (AY547315); PAP from *P. americana* (AY572976); PAP from *Phytolacca acinosa* Roxb. (AY049785); PAP-I from *Phytolacca heterotepala* H. Walt. (AY327475); PAP from *Phytolacca octandra* L. (AF533515); PAP from *P. americana* (AY603352); and PAP-I from *P. acinosa* (AY603353).
PAP-I exhibited a high degree of allelic sequence diversity. The nucleotide sequence identity in the coding region ranged from 80 to 99% with the other isoforms of PAP (Table 1).

Table 1. Sequence identity of PAP I with the other PAP isoforms identified in different Phytolacca spp and locations.

<table>
<thead>
<tr>
<th>Plant species</th>
<th>Location</th>
<th>Accession No</th>
<th>PAP isoform</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. americana</em></td>
<td>China</td>
<td>EF081129</td>
<td>PAP fusion protein</td>
<td>99</td>
</tr>
<tr>
<td><em>P. acinosa</em></td>
<td>China</td>
<td>AY049785</td>
<td>PAP</td>
<td>97</td>
</tr>
<tr>
<td><em>P. heterotepala</em></td>
<td>Italy</td>
<td>AY327475</td>
<td>PAP I</td>
<td>98</td>
</tr>
<tr>
<td><em>P. octandra</em></td>
<td>New Zealand</td>
<td>AF533515</td>
<td>PAP</td>
<td>96</td>
</tr>
<tr>
<td><em>P. americana</em></td>
<td>China</td>
<td>AY603352</td>
<td>PAP-S</td>
<td>89</td>
</tr>
<tr>
<td><em>P. americana</em></td>
<td>China</td>
<td>AY603353</td>
<td>PAPa1</td>
<td>99</td>
</tr>
<tr>
<td><em>P. acinosa</em></td>
<td>China</td>
<td>AY603354</td>
<td>PAPa2</td>
<td>96</td>
</tr>
<tr>
<td><em>P. isularis</em></td>
<td>Korea</td>
<td>AF141331</td>
<td>PIP2</td>
<td>81</td>
</tr>
<tr>
<td><em>P. americana</em></td>
<td>Japan</td>
<td>D10600</td>
<td>PTCA PAP</td>
<td>81</td>
</tr>
<tr>
<td><em>P. americana</em></td>
<td>Japan</td>
<td>AB071854</td>
<td>PAP-S1</td>
<td>80</td>
</tr>
<tr>
<td><em>P. americana</em></td>
<td>Japan</td>
<td>AB071855</td>
<td>PAP-S2</td>
<td>80</td>
</tr>
</tbody>
</table>

The phylogenetic relationship of PAP I with seventeen other PAPs is shown in Figure 5. The PAP-I from Turkey and other PAPs from *P. americana*, grouped into one cluster with 78% bootstrap values. This may be due to the fact that all these RIPs are from the same species of pokeweed only. The alignment of PAP I showed a distant relationship with AB071854 and AB071855 corresponding PAP-S1 and PAP-S2 alleles of *P. americana*.

Figure 5. Phylogenetic relationship of PAP-I to different RIPs. Values given on the nodes are bootstrap values of replications. The nucleic acid sequences of PAP-Turkey|GeneBank (PAP I), (will appear in GenBank as JX446580); PAP fusion protein (EF081129); PAP synthetic construct (EF508129); PAP (X55383); CIP (EU839992); RIP (X96583); RIP (DQ989495); PAP (AY572976); PAP (AY049785); PAP I (AY327475); APP (AF533515); PAP S (AY603352); PAPa1 (AY603353); PAPa2 (AY603354); PIP2 (AF141331); PTCA PAP (D10600); PAP-S1 (AB071854); PAP-S2 (AB071855); and PAP (AY547315) were used.
In this report, we describe the isolation and characterization of a PAP-I gene from the summer leaves of pokeweed plant collected from Trabzon/Turkey. A full-length DNA encoding *P. americana* antiviral protein (PAP) was generated using gene specific forward and reverse primers and the ORF starting at methionine (M) and ending at phenylalanine (F) cloned into pGEM-T Easy vector. We determined the nucleotide sequence of the PAP-I gene and compared the deduced amino acid sequence of other *Phytolacca* plants (Table 1). The full-length PAP sequence was found to be 942 bp long, and has been submitted to the NCBI GenBank (Accession No. JX446580), wherein the deduced antiviral protein has been designated as PAP-I (*P. americana* antiviral protein Type I). The PAP-I DNA contained an open reading frame encoding 313 amino acids, which has the same amino acid number (313 aa) recently reported by Lin et al. (1991).

The deduced amino acid sequence included a putative signal sequence of 22 amino acids at the N-terminus (Fig 2) as reported by Song et al. (2000) for the *P. insularis* antiviral protein (PIP). The first 22 amino acids presumably contain the signal sequence required to translocate the protein to the cell wall, where it has been localized in pokeweed, and is not present in the mature protein (Lin et al. 1991). Dore et al., (1993) reported that a modified PAP gene coding for PAP without the 22-amino acid N-terminal signal peptide failed to express the antiviral protein.

By alignment of PAP-I amino acid sequence with various RIPs, it was revealed that PAP-I contains 14 conserved (AIQMVSEAARFKYI) amino acids. Similar result has been reported by Lin et al. (1991). A multiple alignment analysis of the PAP-I nucleotide sequence with those of other PAPs shows that, on a consensus length of 942 nucleotide residues, 80-99% are identical. According to the phylogenetic analysis and gene structure characteristics discussed above, it could be speculated that the PAP-I gene may be a variant form that resulted from genetic variations such as nucleotide substitutions. Furthermore, no introns have yet been identified in PAP-I gene from other *Phytolacca* species (Poyet and Hoeveler, 1997).

The position of the amino acid alteration in the PAP-I gene was identified by sequence analysis of the plasmids rescued from bacteria. Sequence analysis of PAP-I revealed that it contains a single point mutation, changing the Leucine at position 273 to Phenylalanine (L273F) demonstrated by bidirectional sequencing. However, it is not known that how this divergence, occurring within the gene, affects the N-galactosidase activity of PAP gene. Further studies on currently isolated PAP-I gene are required for a better understanding of its actual N-glycosidase activity, which may affect the level of antiviral activity.

Plant disease management using antiviral proteins of plant origin holds promising to combat viral diseases. Many of these viral inhibitory proteins possess ribosome inactivating property, the most widely accepted mechanism of action as a direct result of their ribosome specific N-glycosidase activity (Rajesh et al. 2005). This work provides some useful information for further understanding the structure and evolutionary relationships within the PAP gene family. Isolation and characterization of the PAP-I gene from *P. americana* may further enrich the gene resources for the obtaining of effective antiviral proteins. The present work, reporting the primary structure of PAP-I gene, may contribute to future biotechnological applications, such as the development of novel therapeutic molecules against plant diseases.

**ACKNOWLEDGEMENTS**

This research was financially supported by a grant from The Scientific and Technical Research Council of Turkey, TUBITAK (Project no: TOVAG-110O180). We are grateful to Prof. V. Pallas Polytechnic University, (Spain) for his continues encouragement and useful suggestions on gene manipulation.

**ÖZET**

Pokeweed antiviral protein (PAP), *Phytolacca americana* L. (Pokeweed) bitkisinde bulunan ve ribosomal RNA’lardaki spesifik bazı purin bazlarını yok ederek ribozomu inaktive etme özelliğine sahip tek zincirli bir proteindir. Sunulan çalışmada pokeweed antiviral protein Tip 1 (PAP-I) geninin tamamı, Trabzon’dan temin edilen *P. americana* bitkisinin yaz yapraklarından

Anahtarkelimeler: Antiviral protein, karakterizasyon, clonlama, PAP-I geni, Phytolacca americana

LITERATURE CITED


Gendron, Y. and Kassanis, B. 1954. The importance of the host species in determining the action of virus inhibitors.


