Molecular Characterization of Soybean Mosaic Virus Nla Protein and its Processing Event in Bacterial Expression

1Bong Kum Choi, 1Jae Sung Park, 1Hye Jin Ahn, 1Hye Jung Yum, 1Vasudevan Ayyappan
2Lakshmi Sumitra Vijayachandran, 1Ji Sue Kim, 1Sei Chang Kim and 1,2Chang Won Choi
1Department of Biology and Medicinal Science, Pai Chai University, Daejeon, Korea 302-735
2Biomedicinal Research Center (RRC), Pai Chai University, Daejeon, Korea 302-735

Abstract: Soybean mosaic virus (SMV)-CN18 is an Rsv resistance-breaking (RB) isolate to overcome soybean resistance genes Rsv1, Rsv3 and Rsv4. The aim of this study was to characterize nuclear inclusion protein a (Nla protein) of RB isolate at the molecular level and demonstrate its processing into genome-linked protein (VPg) and Nla-Pro domains in Esherichia coli containing a bacterial expression pET vector inserted with Nla gene. The full-length of Nla gene was synthesized by reverse transcription-polymerase chain reaction (RT-PCR) and its 1298 nucleotides (nt) and 432 amino acids (aa) were deduced. The nt and aa sequences of Nla gene of SMV-CN18 shared high identities with the corresponding sequences of the Nla gene of the known SMV isolates, suggesting that the Nla is a highly conserved protein. The Nla-Pro domain contains a highly conserved structural motif for proteolysis, while the VPg domain contains a nuclear localization signal (NLS), a putative NTP-binding site and cellular factor-binding sites. The phylogenetic tree revealed that less divergence of Nla protein exists among twelve SMV isolates, which can be supported by a low bootstrap value between clades. In addition, the full-length of Nla gene, amplified by RT-PCR, was ligated into pET-28b E. coli expression vector with an N-terminal His6-tag. Optimal conditions for expression were at 1mM treatment of IPTG at 25°C for 5 hr. The released protein from bacterial lysates remained soluble and proved the processing form of the Nla polyprotein. E. coli expression system shows the processed product of 29 kDa VPg in SDS-PAGE confirmed by western blot analysis in both crude extracts and purified elution products, using Ni2+NTA resin. The present study indicates that the N-terminal region of Nla which is processed and expressed in bacteria.

Key words: Soybean mosaic virus (SMV), Rsv resistance-breaking (RB) isolate, nuclear inclusion protein a (Nla), processing, genome-linked protein (VPg), Nla-Pro, highly conserved protein, E. coli expression

INTRODUCTION

Soybean mosaic virus (SMV), a member of the genus Potyvirus, is a major pathogen of soybean (Glycine max L.). SMV has a positive-sense single-stranded RNA genome of 9588 nucleotides, with VPg on the 5’ end and poly(A) track at the 3’ end. The genome of potyvirus encodes a single large polyprotein that is proteolytically processed by three virus-encoded proteases[13]. One of the proteases, nuclear inclusion protein a (Nia), possesses structural motifs that showed similarity with cellular serine protease, with the substitution of Ser by a Cys as the active site[2,3]. The Nia is actually a 49 kDa polyprotein consisting of two domains, the genome-linked protein (VPg) domain at the N-terminus and the protease (Nia-Pro) domain at the C-terminus[4,5]. The Nia protease plays an important role in the processing of the remaining two-thirds of polyprotein in cis- and trans-processing by catalyzing cleavage of at least six recognition sites[6-9]. The VPg is covalently linked to the 5’ terminal base of a viral genome by a phosphodiester bond[10], needed for virus replication[11]. Its C-terminal region is important for the VPg-VPg self-interaction and its central region is required for the HCpro and VPg interaction[12]. It may be involved in viral RNA synthesis by interacting with the viral RNA-dependent RNA polymerase NiB[13-16], supporting that VPg is a putative primer for potyvirus replication as proposed for picornaviruses[17]. It has been considered as an analog of m’G cap of the mRNAs that might have a role in polyprotein translation because of the its interaction with cap-binding translation initiation factor eIF4E[18,19]. In addition, it is implicated in viral movement from cell to cell[20] or vascular transport[21-23]. Its translocation from inoculated source leaves to the sink leaves, resulted in its accumulation in companion cells at an early stage of infection, suggesting that VPg may be a phloem protein to facilitate virus unloading[24].
Recently, we have reported twelve emerging Rsv resistance-breaking (RB) isolates of SMV, among them SMV-CN18 has an ability to overcome soybean resistance gene Rsv1, Rsv3 and Rsv4, respectively.[25] In this study, the Nla gene was synthesized by reverse transcription (RT)-PCR using SMV-CN18 genomic RNA as a template, compared its sequences of nucleotides (nt) and amino acids (aa) with those of eleven previous reported SMV isolates and characterized its expression in E. coli to observe the processing event into VPg and Nla-Pro domains.

**MATERIALS AND METHODS**

**Viral strain, purification and RNA extraction:** An RB isolate, SMV-CN18, was purified from the infected soybean leaves and its genomic RNA was prepared as described in a previous investigation.[25]

**RT-PCR amplification of Nla gene and cDNA cloning:** The primers used for the amplification of Nla gene were designed based on the conserved nucleotide sequences of known SMV isolates and listed in Table 1. Reverse transcription (RT) reaction was performed on 100 ng of viral RNA in a reaction volume of 20 μl containing 50 mM Tris-HCl, pH 8.3, 75 mM KCl, 10 mM DTT, 3 mM MgCl₂, 1 mM dNTP, 50 pmol of reverse primer, 20 U of RNase inhibitor (Takara, Japan) and 200 U Molony murine leukaemia virus reverse transcriptase (Promega, USA). For amplification, 20 μl of RT mix was added to 80 μl of reaction mixture containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5–4.5 mM MgCl₂, 2.5U Taq polymerase (Takara, Japan), 0.25 mM dNTP, 50 pmol of forward and reverse primers each. The thermocycler (Bio-Rad, Gene Cycler, USA) was programmed for template denaturation at 94°C for 1 min, primer annealing at 55°C for 2 min and 37°C for 3 min. Ten RB isolates of SMV in a previous study.[25] were also subjected to monitor optimal condition for RT-PCR. The purified fragments were directly ligated into pGEM-T Easy Vector System (Promega, USA) and transformed into E. coli JM109 by conventional CaCl₂ procedure. Transformants were selected on Luria-Bertani (LB) medium supplemented with 100 μg mL⁻¹ ampicillin, X-Gal and IPTG, incubating overnight at 37°C.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer polarity</th>
<th>Sequence (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nla-forward</td>
<td>5’ end</td>
<td>ATCAACTCAATGAAAGAAGAG</td>
</tr>
<tr>
<td>Nla-reverse</td>
<td>3’ end</td>
<td>TCTTTTCCTCCTGGACTGTC</td>
</tr>
</tbody>
</table>

aReverse primer containing stop codon (bold, underline), T represents a modified base for termination.
bForward primer containing start codon (bold, underline), T represents a modified base for initiation.

**Expression and purification of the His-tagged recombinant Nla protein:** The pGEM-T Easy vector containing the amplified DNA fragments were digested with NotI, eluted by gel purification and re-ligated into NotI-digested expression vector pET-28b as 6 X His fusion at N-terminus (Novagen, Germany). The goal of this strategy was to put the inserted gene under the control of the IPTG-inducible T7lac promoter. The final constructs were verified by sequencing. Then the recombinant plasmid (pSMV-Nla) was transformed into the expression host E. coli BL21(DE3). Expression of 6 X His-fused proteins was carried out according to manufacturer’s protocol. A single colony was grown in Luria-Bertani (LB) medium containing kanamycin (30 μg mL⁻¹) to an OD₆₀₀ of 0.6, following isopropyl-β-D-1-thiogalactopyranoside (IPTG, final concentration of 0.01-5 mM) was added into the medium and incubated with shaking at 25°C or 37°C for 0-24 h to determine optimal condition. The protein was purified from E. coli cells (100 mL culture) under the native or denaturing conditions (8 M urea) using Ni²⁺-NTA resin (Qiagen, USA) as described in the manual ‘QiAexpressionist’ except that the non-denaturing lysis buffer contained 1 mM lysozyme and 1 mM PMSF. In all cases, suspensions were sonicated three times (30 s each with 1 min interval) prior to stirring at room temperature.

**Nucleotide and amino acid sequences analyses and phylogenetic tree:** The plasmid containing Nla gene was prepared from the transformed bacterial cells using QIA Plasmid Prep Kit (QIAGEN, Germany) and used for sequencing analysis. After linearization of the plasmid containing entire Nla coding region, nt sequences were determined in both directions by the dideoxynucleotide chain termination method using the ABI Model 337 automatic DNA sequencer. The complete genomic sequences of the SMV-CN18 were deposited as EMBL Database Accession no. AJ619757. For sequence comparisons, Nla gene products of the reference isolates were obtained from the NCBI data library including SMV-G2 (S42280) and -G7-1 (AF241739; then designated G7 in USA)[26], SMV-G7d (AY216987)[27], SMV-N (NC002634)[28], SMV-G5b (AY294044; then designated G5 in Korea) and -G7H (AY294045)[29], SMV-HH5 (AJ310200; then designated Huanghuai) and -HZ (AJ312439; then designated Severe)[30], as well as web-published isolates SMV-G72 (AY216010), -Aa (AB100442) and -Aa15-M2 (AB100443). Distance matrices for complete Nla sequences were calculated from the multiple sequence alignments by the DNAMAN version 5.2.9 (Lynnon Biosoft, Quebec, Canada). Phylogenetic analysis was performed using the generated matrices as an input in DNAMAN to build an unrooted tree and the statistical significance of branching was estimated by bootstrap resampled data sets based on 1000 replications.

Table 1: RT-PCR primers used for the cloning of Soybean mosaic virus Nla gene

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer polarity</th>
<th>Sequence (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nla-forward</td>
<td>5’ end</td>
<td>ATCAACTCAATGAAAGAAGAG</td>
</tr>
<tr>
<td>Nla-reverse</td>
<td>3’ end</td>
<td>TCTTTTCCTCCTGGACTGTC</td>
</tr>
</tbody>
</table>
Total proteins were analyzed by 10% SDS-PAGE followed by Western blotting to detect 6X His-tagged protein using Ni\(^{2+}\)-NTA conjugated antibody against 6X His (Qiagen, USA). The membrane was developed by conventional method with nitroblue tetrazolium (NBT)/5-bromo-4-chloro-3-indolylphosphate (BCIP) (MBI Fermentas, Canada) in 100 mM Tris, 100 mM NaCl and 5 mM MgCl\(_2\), pH 9.5.

RESULTS AND DISCUSSION

RT-PCR amplification of NIa gene: Successful amplification of NIa gene from SMV-CN18 was performed by using forward and reverse primers in a range (1.5-4.5 mM) of magnesium concentrations, among which 2.5 mM of MgCl\(_2\) was the optimal condition (Fig. 1a). Under the optimal RT-PCR condition, we detected an array of amplification products of expected size 1,298 bp fragments of NIa gene that encoded 433 aa with a predicted molecular mass of 49 kDa from soybean leaves inoculated with ten SMV RB isolates (Fig. 1b).

Fig. 1: (a) Amplified products of SMV-CN18 NIa gene by RT-PCR under various MgCl\(_2\) concentration. Lanes 1 (1.5 mM), 2 (2.5 mM), 3 (3.5 mM) and 4 (4.5 mM). (b) Amplified products of NIa gene of SMV isolates at 2.5 mM MgCl\(_2\) by RT-PCR. Lanes 1 (CN1), 2 (CN2), 3 (CN3), 4 (CN9), 5 (CN12), 6 (CN13), 7 (CN15), 8 (CN17), 9 (CN21) and 10 (CN31). M represents 1kb DNA ladder

NIa sequence comparisons and phylogenetic analysis: For the investigation of sequence diversity of NIa, the nt and aa sequences of SMV-CN18 were aligned with the corresponding sequences of the NIa gene of the known SMV isolates (data not shown). The highest nt sequence identity was 98% with SMV-HH5, while the lowest nt sequence identity was 92% with that of SMV-N. In aa similarities SMV-CN18 shows the value of 98% with G2 and G5b, respectively and 99% with the remaining nine isolates, suggesting that the NIa is a highly conserved protein among the isolates (Table 2). In addition, it possessed structural motifs for proteolysis that conserved among potyviruses\(^{[8,31]}\). Presumed codons among SMV isolates for the catalytic triad are composed of H\(_{32}\), D\(_{158}\) and C\(_{150}\). On the other hand, the VPg contains a nuclear localization signal (NLS) that is conserved among SMV isolates by the aa residues from 41 to 50 (KKGKKGKSTR). This NLS is quite similar to the NLS (KKGKPKGKTH) in Potato virus A (PVA) VPg\(^{[34]}\) and the NLS (NGGKRGITTR) in Tobacco etch virus (TEV) VPg\(^{[32]}\). SMV VPg also contains a conserved 7 aa residues (A\(_{38}\)YTKKGK) which has been proposed as a putative NTP-binding site in the VPg of PVA, its deletion reduced nucleotide-binding capacity and debilitated uridylylation reaction\(^{[33]}\). Recently, a cellular factor called ‘PVIP’ that interacts with VPg of potyviruses has been identified in some plants. Two domains controlled the interactions with PVIP, suggesting that PVIP plays a role in assistant factor to support potyvirus movement in plants\(^{[34]}\). Sequence comparison of SMV and other four potyviruses, Turnip mosaic virus (TuMV), Lettuce mosaic virus (LMV), PVA and TEV, were aligned for the two domains. It shows that four residues in the first domain (VP aa 1 to 16) and six residues in the second domain (VP aa 40 to 64) are identical with those of all four viruses (Fig. 2).

![Fig. 2: Amino acid sequences of the determinant between VPg and PVIP interaction. SMV VPg sequences were aligned using DNAMAN version 5.2.9 software, with corresponding region from other potyviruses obtained from the NCBI data library. Only regions 1 to 16 and 40 (41 or 42) to 64 (65 or 66) are shown. Identical (star), related (:) and unmatched (triangle) amino acid sequences are indicated.](image)

Table 2: Sequence identities in nucleotides and amino acids of NIa gene between CN18 and known isolates of Soybean mosaic virus

<table>
<thead>
<tr>
<th>Strain/isolate</th>
<th>CN18 Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>93/98</td>
</tr>
<tr>
<td>G5b</td>
<td>93/98</td>
</tr>
<tr>
<td>G7-1</td>
<td>96/99</td>
</tr>
<tr>
<td>G7-2</td>
<td>97/99</td>
</tr>
<tr>
<td>G7D</td>
<td>97/99</td>
</tr>
<tr>
<td>G7H</td>
<td>96/99</td>
</tr>
<tr>
<td>Aa</td>
<td>96/99</td>
</tr>
<tr>
<td>Aa15-M2</td>
<td>96/99</td>
</tr>
<tr>
<td>HH5</td>
<td>98/99</td>
</tr>
<tr>
<td>HZ</td>
<td>97/99</td>
</tr>
<tr>
<td>N</td>
<td>92/99</td>
</tr>
</tbody>
</table>

*Numbers in bold type at right corner refer amino acid sequence identities and numbers in plain type at left corner refer to nucleotide sequence identities.

DNAMAN version 5.2.9 was used in this study to determine the phylogenetic relationships among the NIa proteins of twelve SMV isolates (Fig. 3). In the previous study, the phylogenetic tree revealed that considerable divergence exists among P1 protein of
twenty-five SMV isolates, all those SMV isolates can be grouped into three major types and seven subtypes by similarity clustering\cite{25}. On the other hand, the less divergence of NIa protein existing among twelve SMV isolates, which can be supported by low bootstrap values between clades, suggests that the phylogenetic tree is essentially meaningless. Based on the result, it is logical to hypothesize that the recombination event in the NIa coding region rarely occurred between SMV isolates.

Fig. 3: Phylogenetic analysis for NIa protein amino acid sequences of 12 SMV isolates. The tree (unrooted) was constructed by the DNAMAN version 5.2.9. The grouping occurred after bootstrapping the data (only values > 10% are shown) and the numbers on the branches indicate bootstrap percentages based on 1000 resamplings

Fig. 4: 10% SDS-PAGE of the expressed NIa protein under various concentrations of IPTG at 37°C for 5 h. (a) the supernatant of culture lysate after IPTG treatment, Lanes 1 (wild-type control, 1 mM IPTG treatment), 2 (sample without IPTG treatment) and 3-7 (samples with 0.01 mM, 0.05 mM, 0.1 mM, 0.5 mM and 1mM IPTG treatment, respectively). (b) Lane 1 represents a wild-type control treated with 1 mM IPTG. Lanes 2-6 are samples treated with 1 mM, 2 mM, 3 mM, 4 mM and 5 mM IPTG, respectively. M represents protein Mr marker and unlabelled arrow indicates the VPg

Fig. 5: Time course of NIa protein expression under 1 mM IPTG treatment at 25°C. (a) Culture supernatants and (b) pellets. Lanes 1 (wild type control, culture for 5 h) and 2 (no IPTG treated control, culture for 5 h). Lanes 3-9 represent samples treated with IPTG for 1 h, 3 h, 5 h, 7 h, 9 h, 12 h and 24 h, respectively. M represents protein Mr marker and unlabelled arrow indicates the VPg

Fig. 6: Western blot analysis of the processed NIa protein. The recombinant VPg was detected about size of 29KDa protein using Ni\textsuperscript{2+}-NTA conjugated antibody against 6X His. Lane 1 is crude extract, and lanes 2-7 represent the eluted protein samples during purification using Ni\textsuperscript{2+}-NTA resin

Expression and purification of the His-tagged recombinant NIa protein: The protein was accumulated in high levels at 37°C under 1mM IPTG induction for 5 h but formed inclusion bodies in the bacterial cells (Fig. 4a, b). Optimal condition to express NIa protein as a soluble form in E. coli culture was observed under 1 mM IPTG induction for 5 h at 25°C (Fig. 5a, b). At this temperature, the NIa protein was not only found in the pellets (Fig. 5a) but also in the supernatants of lysates (Fig. 5b). We designed the fusion of six His at the N-terminus of the protein when the coding region of NIa protein was subcloned into pET-28b. Therefore, the VPg with His-tag and NIa-Pro
without His-tag could be released from the processed NLa, and only VPG could be detected by antibody specific to His-tag. Theoretically, the size of native NLa, NLa-Pro and VPg are 49, 27 and 23 kDa, respectively, from the cDNA sequences. Taking into account the N-terminal His-tag, the size of NLa, NLa-Pro and VPg were estimated as 55, 27 and 29 kDa, respectively. The position of band close to the expected size of protein was found at 29 kDa on SDS-PAGE gel (Fig. 5a, b, lanes 3-9), and its migration was corresponding with the predicted VPG. The recombinant protein containing His residues is easily purified by metal-chelation chromatography[35]. To confirm the processing in the purified proteins from E. coli, we performed Western blotting using an Ni2+-NTA conjugated antibody against 6X His. We did not find the unprocessed polyprotein to bind Ni2+ affinity column, but detected the 29 kDa protein corresponding to VPg (Fig. 6). Our results indicate that the anti-His antibodies recognized VPG but failed to recognize unprocessed polyprotein which further processed VPG and NLa-Pro. Our results were corresponding to the previous results that the VPG alone without NLa-Pro was detected when the entire coding region of NLa protein of TuMV was expressed in E. coli[36]. The NLa-Pro of TEV was also expressed in E. coli as a recombinant protein with His-tag, but the expression of unprocessed polyprotein was unsuccessful[37]. The full-length of TuMV NLa gene was expressed as an unprocessed polyprotein (49 kDa) in E. coli by site-directed mutagenesis to block the processing as previously described by Ménard et al.[38]. According to our results and previous reports above mentioned, the unprocessed NLa polyprotein seems to exist for a brief time during its expression in E. coli.

ACKNOWLEDGEMENTS

This work was supported by a grant from Ministry of Commerce, Industry and Energy (MOICE) through the Research Center for Bio-Medicinal Resources (Bio-Med RRC) at Pai Chai University.

REFERENCES


