NUCLEOTIDE SEQUENCE OF THE CAPSID PROTEIN  
GENE OF BEAN YELLOW MOSAIC VIRUS  
CHLOROTIC SPOT STRAIN

Tohru TAKAHASHI, Ichiro UYEDA  
Kazusato OHSHIMA and Eishiro SHIKATA  
Department of Botany, Faculty of Agriculture,  
Hokkaido University, Sapporo 060, Japan  
Received October 31, 1989

Introduction

The potyviruses including bean yellow mosaic virus (BYMV) have a flexuous thread particle of about 750 nm in lengths. The virion contains a single component of (+)sense single-stranded RNA genome of approximately 10K nucleotides. These RNAs are thought to be covalently linked to a small protein (VPg) at the 5’terminal nucleotide and have a polyadenylated track at the 3’terminus.

So far, seven virus-encoded polypeptides were found to be associated with potyviral infections: the coat protein (CP), helper component (HC), VPg, cylindrical inclusion (CI), cell to cell transfor function protein (28–34K) and two nuclear inclusion proteins (NIa and NIb). The cistrons for each of these genes have been mapped to the potyviral genomes. Genomes of potyviruses are expressed through the production of a large polyprotein which is proteolytically cleaved in a similar manner to that of picorna-, toga- and comoviruses.

These studies together with the sequence analyses of the 3’ terminal region of some potyviruses revealed that the predicted coat protein cistron was located upstream from the polyadenylated 3’ terminus of the genome.

Amino acid sequence of the coat protein is an increasingly important taxonomic criterium of the potyvirus as the sequence informations are accumulating and comparison of major potyviruses and their strains become available. It has been pointed out that N-terminus of the amino acid sequence is specific for each potyviruses.

In this paper, we report the nucleotide sequence of the 3’ terminal region of the BYMV-CS. Coat protein gene is mapped in the deduced sequence and its amino acid sequence was determined.

Materials and Methods

Virus and Viral RNA extraction. BYMV-CS was originally isolated from
red clover\textsuperscript{23} and maintained in broad bean (\textit{Vicia fava} cv. Wase-Soramame). The virion was purified from infected broad bean leaves by a procedure described previously.\textsuperscript{39} Genomic RNA from purified BYMV-CS was prepared by a method similar to that described previously.\textsuperscript{4,7,17} To a 2 mg/400 \mu l suspension of the virus was added an equal volume of 200 mM ammonium carbonate (pH 9.0) containing 2 mM EDTA, 2\% SDS, 200 \mu g/ml bentonite and 10 \mu g/ml Proteinase K. After incubation for 20 min at room temperature, the mixture was layered (1 ml/tube) onto a linear-log 7.5\%-33.4\% (w/v) sucrose gradient containing 150 mM NaCl, 15 mM sodium citrate (pH 7.0). Gradients were centrifuged for 9 hr at 24,000 rpm in a Hitachi RPS-27 rotor. Viral RNA, which sedimented as a single sharp zone, was recovered by precipitation with 2.5 volumes of ethanol, and used for cDNA synthesis.

\textit{Complementary DNA synthesis}. The first strand of complementary DNA was synthesized\textsuperscript{13} in a reaction mixture containing 50 mM tris-HCl (pH 7.9 at 42°C), 10 mM MgCl\textsubscript{2}, 30 mM KCl, 1 mM DTT, 1 mM each of dATP, dCTP, dGTP, dTTP, 80 \mu g/ml of oligo (dT\textsubscript{12-18}), 50 \mu g/ml actinomycin D, 100 \mu Ci/ml of (\alpha\textsuperscript{32}P) dCTP, 1600 U/ml reverse transcriptase (BRL) and 200 \mu g/ml of viral RNA and incubated at 42°C for 40 min. After phenol-chloroform extraction and ethanol precipitation, the first strand cDNA was converted to the double stranded cDNA in a reaction mixture containing 100 \mu Ci/ml (\textsuperscript{3}H) dCTP, 20 mM tris-HCl (pH 7.4), 5 mM MgCl\textsubscript{2}, 10 mM ammonium sulfate, 100 mM KCl, 15 \mu M \beta-NAD, 50 \mu g/ml BSA, 10 \mu M dNTPs, 4.3 U/ml RNase H (Takara shuzo Co.), 115 U/ml \textit{E. coli} DNA polymerase I (Takara shuzo Co.), 6 U/ml \textit{E. coli} DNA ligase (BRL). The mixture was incubated sequentially for 1 hr at 12°C and for 1 hr at 22°C. In a second experiment, cDNA was synthesized using a cDNA Synthesis System Plus purchased from Amarsham Corp. After chromatography on Sephadex G-50, the double stranded cDNA was tailed with dCTP using terminal deoxynucleotidyl transferase (Takara shuzo Co.) and annealed with \textit{Pst I} cut g-tailed pBR 322 (BRL) and recombinant cDNAs were used to transform \textit{Escherichia coli} HB 101 cells made competent by CaCl\textsubscript{2}\textsuperscript{28}.

\textit{Southern blot hybridization}. Recombinant DNAs were screened for cDNA inserts by a southern blot hybridization.\textsuperscript{38} Viral RNA in 5 mM MgCl\textsubscript{2} and 25 mM glycine-NaOH (pH 9.0), was incubated at 37°C for 3 hr\textsuperscript{32}, and precipitated by ethanol. Magnesium cleaved viral RNA was labeled using (\gamma\textsuperscript{32}P) ATP and T4 polynucleotide kinase (Takara shuzo Co.) and used as a probe.

\textit{cDNA sequence determination}. Insert cDNAs digested with \textit{Pst I} were ligated into the polylinker region of the replicative form of the bacteriophage M13mp19 and it was used to transform \textit{E. coli} JM109 made competent by RuCl\textsubscript{2}\textsuperscript{15,19}. An ordered set of deleted clones was made by exonuclease-III attacking method\textsuperscript{19} and sequenced by dideoxy chain termination method using a sequencing kit (Takara shuzo Co.).
Sequences were compiled and analysed using a NEC 9801 XL computer and a software programmed by GENETYX (SDC Software Development Co.).

Characterization and amino acid sequencing of the capsid protein. Immediately after the purification, 1 mg of purified virus was resuspended in 2 ml of 50 mM sodium borate buffer at pH 8.0, and incubated with 2 μg of trypsin (Sequence grade, Boehringer Mannheim) for 30 min at room temperature. The trypsin-resistant particles which contain core subunits was collected by centrifugation at 40,000 rpm for 2 hr at 4°C in Hitachi RP-65 rotor. The final pellet was resuspended in double distilled water. SDS-PAGE of BYMV-CS intact or core capsid proteins was carried out in 12.5% polyacrylamide gel electrophoresis system of LAEMMLI. Amino acid composition of BYMV-CS intact coat protein was determined by using hydrolysate treated in 6N HCl at 110°C for 24 hr, and the amino acid analysis was performed by a Hitachi-835 analyzer. The N-terminal sequence of intact or trypsin treated virion were performed by a sequential Edman degradation method in 477A Protein Sequencer and 120A PTH Analyzer (Applied Biosystems).

Results

Recombinant DNA Molecules Containing BYMV-CS

About four hundreds of recombinant DNA plasmids containing cDNA inserts derived from BYMV-CS RNA were identified. The cDNA inserts of those plasmids ranged in size from 500 to 2,200 base pairs (bp), and were analyzed by

![Fig. 1. Restriction analysis and southern blotteds of DNA from the clones. Plasmid DNAs containing cDNA inserts were restricted with PstI and electrophoresised in 1.5% agarose gel(A). DNAs were transferred onto nitrocellulose, hybridized with 32P labeled genomic RNA and autoradiographed(B).](image-url)
Southern blot hybridization with 5' 32P-labeld genomic RNA probe as shown in Fig. 1. In this study, we selected 6 recombinant cDNA clones which covered 3' terminal portion (pCS35, pCS67, pCS84, pCS119, pCS125, pCS2-50,) of the genome RNA. The relationships between the cDNA inserts and the BYMV-CS genomic RNA are presented in Fig. 2, along with a strategy used to determine the nucleotide sequence (Fig. 3). Since these clones did not contain the 3' terminus as they lacked the poly A sequence up stream of the oligo dC tail, the second cDNA synthesis was performed using a kit purchased from Amasham Corp. The plasmid pCS201 was found to contain the poly A sequence.

**DNA Sequencing**

The 1,392 nucleotides sequence determined from pCS119 and pCS201 was presented in Fig. 4. The DNA sequence contains one large open reading frame of 1,218 nucleotides which is terminated by a single TAG codon (Fig. 4.). The

**Fig. 2.** The physical maps and asignment of complementary DNA recombinant molecules to BYMV-CS genomic RNA.

**Fig. 3.** The nucleotide sequencing strategy of 3'terminal region of BYMV-CS. The horizontal arrows indicated the direction and extent of sequence determination. Since pCS119 did not contained poly A sequence, poly A and upstream sequences were determined using a fragment of pCS201 digested with PstI and XhoI.
Fig. 4. The cDNA sequence of BYMV-CS as the genomic RNA sense. The predicted amino acid sequence of the large open reading frame is presented in the standard three letter code above the nucleotide sequence. The cleavage site is indicated by black arrow.
polypeptide encoded for is 406 residues long (Fig. 4.). The other reading frames contain numerous stop codons.

*Coat protein analysis*

BYMV-CS coat protein has a molecular weight of about 30,808 which was estimated from amino acid composition analysis (Table 1). This value is in good agreement with that estimated from SDS-PAGE analysis (Fig. 5). An attempt to determine the amino acid sequence of the N-terminal portion of purified BYMV-CS coat protein was unsuccessful, presumably due to a blocked N-terminus. Similar situations are also encountered with those of TEV-HAT, TVMV and PVY.2,6,5).

But amino acid sequence of these viruses were obtained after removing the N-terminal about 20 to 30 amino acid residues by partial digestion with trypsin. When purified BYMV-CS was likewise treated with trypsin, the intact 32K coat protein resulted in the generation of a "trypsin resistant core" with an estimated molecular weight of 29K as determined by SDS-PAGE (Fig. 5). The sequence of

<table>
<thead>
<tr>
<th>Amino acid residue</th>
<th>BYMV-CS Predicted NA Seq.</th>
<th>Determined chemically</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Asx</td>
<td>37</td>
<td>36</td>
</tr>
<tr>
<td>Thr</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>Ser</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Glx</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>Gly</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>Ala</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>Cys</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Val</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Ile</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Leu</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>Tyr</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Phe</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Lys</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>His</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Arg</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Pro</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Trp</td>
<td>3</td>
<td>N</td>
</tr>
<tr>
<td>mol. wt.</td>
<td>30,998</td>
<td>30,808</td>
</tr>
</tbody>
</table>

**Fig. 5.** Polyacrylamide gel electrophoresis of intact (B) and trypsin treated (A) coat protein of BYMV-CS. Numbers on the right are Mw of marker proteins (C)
the N-terminal portion of this trypsin-resistant core was determined chemically to be Q-I-I-P-D-R-T-V-N-A. This sequence was identical to amino acid residues -237 through -228 from the carboxyl terminus predicted from the nucleotide sequence (Fig. 6). From these results, the coat protein gene was identified as shown in Fig. 4. Comparison of the amino acid sequence of BYMV-CS coat protein with other potyviruses showed that the extents of amino acid sequence homology with other potyvirus coat proteins are 92%, 64%, 53%, 59%, 59% and 53% for BYMV-GDD, TEV, TVMV, PeMV, PVY and SCMV, respectively.3,6,8,12,14,34).

Discussion

Cell free translation studies of some potyviruses indicated that the genome was generated as one large protein which was subsequently processed to give the mature proteins.5,7,18,40). Similar gene expression mechanism was demonstrated for picornaviridae,22) and subsequently been found for cowpea mosaic virus42) or other plant viruses.43). It was shown that the enzymes involved were cystein like proteinase and catalysis occurred on the carbonyl side of a glutaminyl residue. In potyviruses, the protein cleavages occurred on Gln-Gly, Gln-Ser or Gln-Ala,1,2,3,6,8,12,27,31), and amino acids with "bulky" side chains are found adjacent to the glutamine residue and these amino acids are tyrosine, phenylalanine or histidine of aromatic amino acids.1) In our amino acid sequence predicted from the nucleotide sequence, these glutamine-X dipeptide were found at five positions. The molecular weight of the BYMV-CS protein was estimated to be 32K by polyacrylamide gel electrophoresis. The value suggests that the probable cleavage site is between the glutamine and serine residues found at amino acid position -273 to -272. This would result in a 30,998 molecular weight capsid protein molecule, which is in good agreement with the molecular size of 30,808 estimated by amino acid composition analysis and also 32K estimated by polyacrylamide gel electrophoresis. In addition, chemical sequencing of the protein digested with trypsin yielded amino terminal protein of ten amino acids which was identical to amino acid residues -237 to -228 of that deduced from the nucleotide sequence. This ten amino acid sequence was not found at any other locations in the putative
polyprotein. Tryptic cleavage after the lysine residue at -238 of the predicted coat protein would yield a trypsin resistant core protein with molecular weight of 27,479 similar to that estimated from polyacrylamide gel electrophoresis. Thus, we concluded that coat protein is produced by proteolytic cleavage of BYMV-CS polyprotein between the glutaminyl residue at amino acid position -274 and the following seryl residue. The same dipeptide was apparently cleaved in TEV -HAT and TVMV. Alignment of amino acid sequences showed that BYMV-CS coat protein has significant homologies with other potyvirus coat proteins. The inner portion of amino acid sequences have higher homologies than that of N- and C-terminal sequences between different viruses (Fig. 7). Comparison of the amino acid sequence between BYMV-CS and BYMV-GDD showed that extensive homology existed even in N- and C-terminal regions. These results were also consistent with the notion that immunological specificities of potyviruses were located in N- and C-terminal arms arranged on the surface of virus particle as proposed previously. 20,36,37,38).

Nucleotide sequence of the 3'noncoding sequence of BYMV-CS has no homology to that of other potyviruses sequenced to data and no polyadenylation signals were found. However significant homology in the 3'noncoding region between BYMV-CS and BYMV-GDD was found.

![Fig. 7. Harr Plot analysis of six potyvirus(vertical) coat proteins with BYMV-CS(horizontal). The dot-plots were made using a window of 10 and stringency of 0.7.](image)

**Summary**

The sequence of 1,392 nucleotides of the 3'region of the bean yellow mosaic virus chlorotic spot strain (BYMV-CS) RNA has been determined. There is a single open reading frame (ORF), from the 5'end, of 1,218 nucleotides upstream from a 173 nucleotides non-coding region that is polyadenylated. The ORF encodes a polypeptide of 406 amino acids. The cistron of the genome encoding a
potyviral capsid protein is adjacent to the 3'non-coating region. The chemically
determined amino acid composition of the coat protein was similar to that
predicted from the nucleotide sequence. The sequence of the 10 amino terminal
amino acids of the BYMV-CS coat protein was determined chemically after
removing blocked amino terminal oligo-peptides. Our data suggeste that the
BYMV-CS coat protein, like that of other potyviruses, is a product of the
maturation of a large polyprotein. The putative cleavage site is at a glutamine
-serin dipeptide. The capsid protein gene consists of 819 nucleotides and corre­
sponds to a region coding 273 amino acids which have a calculated molecular
weight (Mw.) of 30,998. The amino acid sequence of BYMV-CS coat protein is 53
to 64% homologous to other potyvirus coat protiens and most of the variations
are found in the amino terminal region. And it was 92% identical to coat protein
gene of BYMV-GDD.

Key words: BYMV, Capsid protein, Nucleotide Sequence.

Acknowledgment

This work was supported in part by a Grant in Aid for Scientific Reseach
(No. 63790330) from the Ministry of Education, Science and Culture, Japan and
Legume Association (Mame-Rui Kikin Kyokai) and a grant from Akiyama
Foundation.

The authors thank Dr. K. Fujinaga, Sapporo Medical College, for his advice
on cDNA cloning and sequencing techniques.

Literature Cited

1) ALLISON, R. F., SORENSON, J. C., KELLY, M. E., ARMSTRONG, F. B. and DOUGHERTY, W. G.: 
Sequence determination of the capsid protein gene and flanking region of tobacco etch 
virus: Evidence for synthesis and processing of a polyprotein in potyvirus genome 
2) ALLISON, R. F., DOUGHERTY, W. G., PARKS, T. D., WILLIS, L., JOHNSTON, R. E., KELLY, M. 
and ARMSTRONG, F. B.: Biochemical analysis of the capsid protein gene and capsid 
protein of tobacco etch virus: N-terminal amino acids are located on the virion's 
surface. Virology 147: 309-316. 1985
3) ALLISON, R., JOHNSTON, R. E. and DOUGHERTY, W. G.: The nucleotide sequence of the 
coating region of tobacco etch virus genomic RNA: Evidence for the synthesis of a 
4) BRAKKE, M. K. and van PELT, N.: Properties of infectious ribonucleic acid from wheat 
yellow mosaic virus RNA: Inhibition of proteolytic processing by antiserum to the 49K 
6) DOMIER, L. L., FRANKLIN, K. M., SHAHAKUDDIN, M., HELLMANN, G. M., OVERMEYER, J. H.,


