Arch Virol (1995) 140: 1473–1482

Archives of Virology

Nucleotide sequence analysis of two nuclear inclusion body and coat protein genes of a sweet potato feathery mottle virus severe strain (SPFMV-S) genomic RNA

Brief Report

M. Mori[1,*], J. Sakai[1], T. Kimura[1], T. Usugi[2], T. Hayashi[1,], K. Hanada[1], and M. Nishiguchi[1,*]**

[1]Kyushu National Agricultural Experiment Station, Kumamoto, [2]Japan International Research Center for Agricultural Sciences, Okinawa Branch, Okinawa, Japan

Accepted April 7, 1995

Summary. Recombinant DNA molecules containing cDNA of a sweet potato feathery mottle virus severe strain (SPFMV-S) RNA genome were constructed and the partial nucleotide sequences were determined for three DNA inserts, which cover 4.2 kb from the 3′-terminus excluding the poly (A) tail. This region of the genome consists of an open reading frame of 1340 amino acids (a.a.) and a 3′-non-translated region of 224 nucleotides. The protein products expected were $6K_2$ (53 a.a.), NIa (435 a.a.), NIb (521 a.a.) and CP (315 a.a.). Among NIa, NIb and coat proteins, the NIb protein was found to be the most conserved (59–68%) when compared to the corresponding proteins of other distinct potyviruses.

*

Sweet potato feathery mottle virus (SPFMV) is a member of the potyvirus group and is transmitted by aphids in a non-persistent manner [16]. SPFMV is widespread throughout the world in cultivated areas of sweet potato [18]. Usugi et al. [28] reported three viruses from diseased sweet potato in Japan including SPFMV, sweet potato latent virus, and sweet potato symptomless virus. The SPFMV isolate was referred to as an ordinary strain of SPFMV (SPFMV-O), and caused slight damage to the sweet potato. They further isolated another strain of SPFMV, designated as SPFMV-S, which causes "Obizyo-sohi" disease on the fleshy roots of sweet potato [29] which is similar to the russet crack disease

Present addresses: *National Institute of Agrobiological Resources, **Japan International Research Center for Agricultural Sciences, Tsukuba, Japan.

previously reported [7]. The virus particles of SPFMV-O and -S are 850–880 nm long and SPFMV-S is serologically discriminated from SPFMV-O as well as from SPFMV-RC [29].

Potyviruses have a positive sense, ss RNA genome which encodes a large polyprotein. After translation, the polyprotein is cleaved into a set of mature proteins including coat, nuclear inclusion body a (NIa) and b (NIb), $6K_2$, cytoplasmic inclusion body (CI), $6K_1$, P3, helper component protease (HC-Pro) and P1 proteins [23]. The coat protein gene of plant viruses including potyviruses has been used to produce transgenic plants resistant to viruses [5].

As a step toward genetically engineering virus resistance in sweet potato as well as a contribution toward efficient diagnosis of this virus, we have begun to elucidate the genome organization of SPFMV-S RNA. The 3′-terminal 2.3 kb nucleotides of the SPFMV-O genomic RNA [15], and the coat protein gene and 3′-non-coding region of two other strains (RC and C) of SPFMV [1] have been reported. In this paper, we present the sequence of 4.2 kb nucleotides from the 3′-terminus of SPFMV-S RNA and compare the deduced amino acid sequences of the encoded proteins including $6K_2$, NIa, NIb and CP with those of other SPFMV strains and other potyviruses.

Virus particles of SPFMV-S were purified from infected *Ipomoea nil* leaves (60–100 g) which had been stored at −85 °C as previously described [28]. The genomic RNA was prepared by a previously described method [17] with the following slight modification. The RNA was prepared by phenol extraction followed by ethanol precipitation instead of sucrose gradient ultracentrifugation.

Two μg of RNA was used for cDNA synthesis. cDNA was synthesized by the method of Gubler and Hoffman [10] using the cDNA Synthesis Kit (Pharmacia-LKB) and oligo-$(dT)_{12-18}$ primer, and connected to the CIP (alkaline phosphatase from calf intestine) treated *Eco*RI site of pBluescript II SK+ (Stratagene). *E. coli* (HB101) was transformed with the ligated mixture. Transformants were used for colony hybridization with a probe of ^{32}P-labeled 2.3 kb *Eco*RI fragment, which contains the 3′-terminal region of the cDNA of SPFMV-O [15]. Plasmid DNAs were purified by the rapid preparation method [26], digested with *Eco*RI and electrophoresed. pVC1 and pVC2 carried DNA inserts, 1.8 kb and 3.5 kb, respectively. The 3.5 kb insert was divided into two *Eco*RI fragments, 2.3 kb and 1.2 kb. Each fragment was subcloned into the *Eco*RI site of pBluescript II SK+ and designated pVC2a and pVC2b (Fig. 1). To further clone the upstream region of the genome, oligo nucleotides PR5 (5′-ACAAACTCCCCATGACG-3′), which is complementary to the 961 bp to 977 bp in Fig. 2, was synthesized using the model 391 PCR-mate DNA synthesizer (ABI) and used as a primer for cDNA synthesis. This primer and TimeSaver cDNA Synthesis Kit (Pharmacia-LKB) were used for cDNA synthesis. As a result, pVC3 containing 1.0 kb cDNA was obtained (Fig. 1).

DNA sequencing was conducted by the dideoxynucleotide chain terminator method [25] using the 373A DNA sequencer (ABI) and Dye Primer Cycle Sequencing Kit (ABI). Double stranded DNA templates were generated from the ordered deletion plasmids constructed by the Erase-a-base system

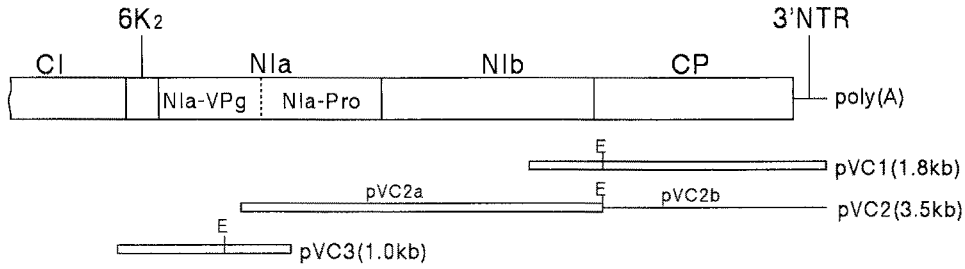

Fig. 1. A tentative map of the 3′-terminal region of SPFMV-S. Vertical lines indicate the proposed cleavage sites for NIa-Pro. The vertical dashed line indicates the proposed internal cleavage site in the NIa. 3′-NTR indicates the 3′-non-translated region. Also shown is a map of the SPFMV-S cDNA clones used for subcloning and sequencing. Square represents the sequenced region. The positions of the *Eco*RI sites are indicated *(E)*. *CP* Coat protein

(Promega). All sequences were determined in both directions. Sequence data were analyzed with software from DNASIS (Hitachi, Japan) or GENETYX (SDC, Japan).

pVC1, pVC2a and pVC3 were selected for sequencing. The sequence of the 3′-terminal 4244 nucleotides is shown in Fig. 2 together with the unique large open reading frame (ORF) encoding 1340 amino acids (excluding the stop codon). The 3′-non-translated region (NTR) is 224 nucleotides excluding the poly (A) sequence, which is as long as that of the SPFMV-O strain [15], however, two bases longer than those of RC and C strains [1]. pVC1 contains 20 adenine residues at the 3′-terminus.

The 3′-NTR of SPFMV-S is homologous to the corresponding regions of SPFMV-O [15], RC, and C strains [1], showing 97.7%, 97.3%, and 98.2% identity, respectively. The nucleotide sequence necessary for the unique 3′-terminal secondary structure as previously described [15] is completely conserved and shown by the horizontal arrows in Fig. 2.

By analogy to the proteinase cleavage sites of other potyviruses [23], the cleavage sites by NIa-proteinase were predicted. Each of the proposed SPFMV-S NIa proteinase cleavage sites occurs between glutamine (Q) and either threonine (T), serine (S), or glycine (G). Table 1 shows the conserved amino acid sequence around each of the four proposed SPFMV-S NIa proteinase cleavage sites, as well as the deduced consensus cleavage site sequence and the NIa internal cleavage site sequence. The resulting CP, NIb, NIa and $6K_2$ are 315, 521, 435, and 53 a.a. in length, respectively.

The identity of the SPFMV-S proteins with those of the other potyviruses is listed in Table 2. It appeared that the NIb region is highly conserved (59%–68%) and that PPV and TuMV are the ones most closely related to SPFMV-S. The close relationship between PPV and SPFMV in the coat protein region has been reported previously [1].

The amino acid sequence of CP has been predicted for three strains [1, 15]. A multiple alignment of the amino acid sequence of the CP is shown in Fig. 3. SPFMV-S CP shares 97%, 98% and 84% identity with those of O, RC and C

```
CGG GCG TAC GTA CGT GAG ACC GGA GCC ACC TCT TGT GTG CTT CAT CAG ACA AAA GAT TCT CTC TCA AAG CAC      72
 R   A   Y   V   R   E   T   G   A   T   S   C   V   L   H   Q   T   K   D   S   L   S   K   H      24

CTT CGT TTG AAA GGA GTG TGG AAT AAA TCA GTC ATC ACA CAA GAT TTG TTC ATT TTA GCA GGG GTA TTT GCT     144
 L   R   L   K   G   V   W   N   K   S   V   I   T   Q   D   L   F   I   L   A   G   V   F   A      48

GGT GGC TTA TGG ATG ATA ATG GCA GGG CTA AAG GAG TCA TTT GAT CAA ACT GTG CAA CAT CAA GGA AGA GAA     216
 G   G   L   W   M   I   M   A   G   L   K   E   S   F   D   Q   T   V   Q   H   Q   G   R   E      72

AAG AGA CAG ATG CAG AAA TTG AAA TTC AGA AAA GCG CGT GAT AAT AAA TTG GGA TTC GAG GTG CAC GCT GAT     288
 K   R   Q   M   Q   K   L   K   F   R   K   A   R   D   N   K   L   G   F   E   V   H   A   D      96

GAT GGT ACA ATC GAG CAC TTC TTT GGT AGT GCG TAT ACC AAG AAA GGG AAG CAG AAG GGT AAG GTC ACC GGA     360
 D   G   T   I   E   H   F   F   G   S   A   Y   T   K   K   G   K   Q   K   G   K   V   T   G     120

ATG GGA TCA AAG AAT AGA AAA TTT ATA AAT ATG TAT GGA TTT GAT CCC ACT GAA TAC TCA TTT GTG CGG TTC     432
 M   G   S   K   N   R   K   F   I   N   M   Y   G   F   D   P   T   E   Y   S   F   V   R   F     144

GTT GAC CCA TTG ACC GGA GCT GTG ATA GAT GAC TCT CCA TAT ACC GAT ATC CTC CTT GTT CAA GAA AGG ATT     504
 V   D   P   L   T   G   A   V   I   D   D   S   P   Y   T   D   I   L   L   V   Q   E   R   I     168

GGT GAG GCT CGT CTG AAT GCT ATT AAA GAA GAT GAA CTA TCA AGA GAA AAA GTA GCA CAG AAC CCA GGA ATT     576
 G   E   A   R   L   N   A   I   K   E   D   E   L   S   R   E   K   V   A   Q   N   P   G   I     192

CAT GCA TAT TAC ATT AAT GAA ATC ACA AAT GCT GCG CTC AAA GTT GAT CTC ACA CCA CAC AAC CCA CTA CTT     648
 H   A   Y   Y   I   N   E   I   T   N   A   A   L   K   V   D   L   T   P   H   N   P   L   L     216

GCA TGT GAG AGG CAT AGT ACT ATA GCT GGT TAT CCT GAA TAT GAA GGG GTA CTG AGA CAA ACT GGG CAC CCT     720
 A   C   E   R   H   S   T   I   A   G   Y   P   E   Y   E   G   V   L   R   Q   T   G   H   P     240

ATT AAG ATG ACG CTG AAT GAC GTT CCG AAA AGT CCT GAG GAA ACA AGT CTA GTA GGT CAT GAA AGC AAA TCA     792
 I   K   M   T   L   N   D   V   P   K   S   P   E   E   T   S   L   V   G   H   E   S   K   S     264

CTA TTT AGA GGT CTG AGG GAT TAT AAT CCG ATA GCT AGC GTC ATA TGC CAT CTC GTG AAT GAG GCG GAT GGA     864
 L   F   R   G   L   R   D   Y   N   P   I   A   S   V   I   C   H   L   V   N   E   A   D   G     288

CGA ACA AGT GAT TGT TTC GGA ATT GGA TAT GGT GGT CTT ATT GTC ACC AAT AGG CAT CTG TTT AAA CGA AAC     936
 R   T   S   D   C   F   G   I   G   Y   G   G   L   I   V   T   N   R   H   L   F   K   R   N     312

AAT GGA ACG TTG ACA ATA AGG TCT CGT CAT GGG GAG TTT GTC ATT AAG AAC ACA ACT CAA CTT GGG ATG AAG    1008
 N   G   T   L   T   I   R   S   R   H   G   E   F   V   I   K   N   T   T   Q   L   G   M   K     336

CCC TGT GCT GAT AGA GAC ATA CTC ATC ATC AGG ATG CCT AAA GAT ATA CCC CCG TTC CCA CAA CGC CTG AAA    1080
 P   C   A   D   R   D   I   L   I   I   R   M   P   K   D   I   P   P   F   P   Q   R   L   K     360

TTT AGG GTT CCA AAG GAA AAT GAG AGA ATC TGC TTG GTT GGG TCA AAC TTT CAG GAC AAA TCA ATC ACA AGC    1152
 F   R   V   P   K   E   N   E   R   I   C   L   V   G   S   N   F   Q   D   K   S   I   T   S     384

ACG ATT TCT GAA ACC AGT GTG ACG TGT CAT GTA CCA AAC TCG CAT TTT TGG AAG CAC TGG ATT GAC ACA AAG    1224
 T   I   S   E   T   S   V   T   C   H   V   P   N   S   H   F   W   K   H   W   I   D   T   K     408

GAT GGG CAT TGT GGA CTT CCC TTA GTA AGC ACT ACC GAT GGT GCT TTG TTG GGA GTA CAT AGT TTG TCA AAT    1296
 D   G   H   C   G   L   P   L   V   S   T   T   D   G   A   L   L   G   V   H   S   L   S   N     432

TTA ACG AAT ACC CAA AAT TTC TTC GCC TCA TTT CCT GAG AAC TTT GAA GTA GAT TAT TTG AAA ACT CCC GAG    1368
 L   T   N   T   Q   N   F   F   A   S   F   P   E   N   F   E   V   D   Y   L   K   T   P   E     456

GCA ATG GAT TGG ATA AAG AAA TGG AGT TAT AAT CCA GAT GAA ATT TGT TGG GGA ACG CTG GAG TTA AAA ACA    1440
 A   M   D   W   I   K   K   W   S   Y   N   P   D   E   I   C   W   G   T   L   E   L   K   T     480

GGA CAA CCC ATT GCT CCA TTT AAA GTG TCT AAG CTG ATC ACA GAC CTA GAA GGT ATA CAA GTG TAT GCT CAA    1512
 G   Q   P   I   A   P   F   K   V   S   K   L   I   T   D   L   E   G   I   Q   V   Y   A   Q     504

ACA AGG TCG GAT AGA TGG GTT CAA GAT AGA TTA TAT GGC AAT CTC AAG GCA GTT GGT CAA TGT CCT GCA CAA    1584
 T   R   S   D   R   W   V   Q   D   R   L   Y   G   N   L   K   A   V   G   Q   C   P   A   Q     528

TTG GTT ACC AAG CAT GTT GTG AAA GGG AAA TGC ATG CTA TTT GAC CTT TAT TTG CAG CAA GAT CAA TCT GAG    1656
 L   V   T   K   H   V   V   K   G   K   C   M   L   F   D   L   Y   L   Q   Q   D   Q   S   E     552

AAG GAG TAT TTT AAG CCT TTA ATG GGA GCA TAT GGG AAA AGT AGG CTC AAC AAA GAG GCA TAC AAC AAG GAT    1728
 K   E   Y   F   K   P   L   M   G   A   Y   G   K   S   R   L   N   K   E   A   Y   N   K   D     576

TTG TTT AAG TAT GCG ACA CAA ATT CAG GCA GGA GAT GTG CAA GTA GAT ATG TTC GAA TTG GCG GAA AGA TCT    1800
 L   F   K   Y   A   T   Q   I   Q   A   G   D   V   Q   V   D   M   F   E   L   A   E   R   S     600

GTC GTT TCT ATG CTT ACA GCA AAA GGA TTT GAA AAA TGC AAT TAC ATC ACG GAT CCA GAG GAG ATT CTG AAA    1872
 V   V   S   M   L   T   A   K   G   F   E   K   C   N   Y   I   T   D   P   E   E   I   L   K     624

GCA CTG AAT ATG AAA GCC GCG GTT GGC GCG ATG TAC AGT GGG AAA AAG AAG GAT TAC TTT GAA GGG ATG AGT    1944
 A   L   N   M   K   A   A   V   G   A   M   Y   S   G   K   K   K   D   Y   F   E   G   M   S     648

GAC CAT GAT GTA GAA GAC CAT CTT TTT CAC AGT TGC AAA CGC TTG TTT ATG GGA TAC AAA GGT CTT TGG AAC    2016
 D   H   D   V   E   D   H   L   F   H   S   C   K   R   L   F   M   G   Y   K   G   L   W   N     672

GGG TCC TTG AAA GCT GAG CTG AGG CCT ATG GAA AAA GTG GAA CTC AAC AAA ACA AGA ACT TTC ACA GCG GCC    2088
 G   S   L   K   A   E   L   R   P   M   E   K   V   E   L   N   K   T   R   T   F   T   A   A     696

CCT CTT GAC ACG CTA CTT GGT GGC AAG GTG TGT GTG GAT GAT TTT AAT AAC ATG TTC TAC AAC CAC CAT TTA    2160
 P   L   D   T   L   L   G   G   K   V   C   V   D   D   F   N   N   M   F   Y   N   H   H   L     720

AAG TGT CCA TGG ACG GTG GGA ATA ACT AAA TTC TAT CAA GGA TGG GAT AGA CTA CTC ACA TCT CTA CCT GAG    2232
 K   C   P   W   T   V   G   I   T   K   F   Y   Q   G   W   D   R   L   L   T   S   L   P   E     744

GGA TGG ATC TAT TGT GAC GCC GAC GGC TCG CAA TTT GAC AGC TCT CTT TCT CCA TAC TTG ATT AAT TCA GTT    2304
 G   W   I   Y   C   D   A   D   G   S   Q   F   D   S   S   L   S   P   Y   L   I   N   S   V     768

CTT AAT ATA AGA AGA GAA TTT ATG GAA GAT TGG GAT GTC GGT GAT CAA ATG CTC CGG AAT TTA TAC ACA GAG    2376
 L   N   I   R   R   E   F   M   E   D   W   D   V   G   D   Q   M   L   R   N   L   Y   T   E     792
```

```
ATA GTA TAC ACA CCA ATA CTG ACA CCA GAC GGA ACT ATA GTG AAG AAA TTC AAA GGT AAC AAC AGT GGA CAA   2448
 I   V   Y   T   P   I   L   T   P   D   G   T   I   V   K   K   F   K   G   N   N   S   G   Q    816

CCA TCC ACT GTG GTT GAT AAC ACA CTT ATG GTT GTG TTA GCA GTA CAC TAC ACA CTG CTA AAA CTA GGT ATT   2520
 P   S   T   V   V   D   N   T   L   M   V   V   L   A   V   H   Y   T   L   L   K   L   G   I    840

CAG GAG AGC GAA TTT GAT GAA TGC TGC GTA TTC TTC GCG AAT GGA GAT GAT CTA TTA CTG GCA ATG AGG CCA   2592
 Q   E   S   E   F   D   E   C   C   V   F   F   A   N   G   D   D   L   L   L   A   M   R   P    864

GAC ACA GCT CAT TTA CTG GAT AAG TTT AGT GAG TGC TTT TCA GAG TTA GGA CTC AAT TAT GAT TTT TCA TCG   2664
 D   T   A   H   L   L   D   K   F   S   E   C   F   S   E   L   G   L   N   Y   D   F   S   S    888

CGA ACC AGT AAT AAA GAA GAG CTA TGG TTT ATG TCA CAC CGC GGA TTG AAA CGT GAT GGA ATA TTC ATA CCG   2736
 R   T   S   N   K   E   E   L   W   F   M   S   H   R   G   L   K   R   D   G   I   F   I   P    912

AAG CTG GAG CCT GAG AGA ATT GTT TCC ATT CTT GAA TGG GAT CGC TCA CAC GAA CCG ATT CAT CGA TTG GAA   2808
 K   L   E   P   E   R   I   V   S   I   L   E   W   D   R   S   H   E   P   I   H   R   L   E    936

GCA ATA TGT GCT GCG ATG GTA GAA TCA TGG GGT TAT GAT GAA CTT CTT CAT CAT ATC AGA AAA TTC TAT GCA   2880
 A   I   C   A   A   M   V   E   S   W   G   Y   D   E   L   L   H   H   I   R   K   F   Y   A    960

TGG GTA TTG GAC CAA GCC CCA TAC AAT GAA TTA GCA CGA AGT GGG AAG GCA CCA TAC ATA GCT GAA ACT GCA   2952
 W   V   L   D   Q   A   P   Y   N   E   L   A   R   S   G   K   A   P   Y   I   A   E   T   A    984

CTT AAA GCA CTC TAT ACT GGT GTC CAA CCT AGT GCT TCT GAA TTG AGT GCG TAT GCA AAA GTG CTT AAC GAA   3024
 L   K   A   L   Y   T   G   V   Q   P   S   A   S   E   L   S   A   Y   A   K   V   L   N   E   1008

ATG TAT GAT GAT AGT GTA CTT CAG GAG AAT GAG TTA GAA GTA TAC CAT CAA TCT AGT GAA CGT ACT GAA TTC   3096
 M   Y   D   D   S   V   L   Q   E   N   E   L   E   V   Y   H   Q   S   S   E   R   T   E   F   1032

AAA GAT GCG GGA GCG AAC CCT CCA GCC CCT AAG CCT CAG AAT ATC CCT CCA CCA CCC ACA ATA ACT GAG GTT   3168
 K   D   A   G   A   N   P   P   A   P   K   P   Q   N   I   P   P   P   P   T   I   T   E   V   1056

ACT GAT CCA GAA GAC CCA AAG CAG GCA GCT TTG AGA GCT GCA CGA GCT AAG CAA CCC GCA ACC ATT CCA GAA   3240
 T   D   P   E   D   P   K   Q   A   A   L   R   A   A   R   A   K   Q   P   A   T   I   P   E   1080

TCA TAT GGA CGA GAC ACT AGC AAG GAG AAG GAA TCA ATA GTG GGG GCA TCA TCA AAG GGT GCG AGG GAT AAA   3312
 S   Y   G   R   D   T   S   K   E   K   E   S   I   V   G   A   S   S   K   G   A   R   D   K   1104

GAT GTA AAC GTT GGA ACA GTT GGT ACG TTT GTC GTG CCA CGT GTT AAG ATG AAT GCA AAC AAG AAA AGG CAA   3384
 D   V   N   V   G   T   V   G   T   F   V   V   P   R   V   K   M   N   A   N   K   K   R   Q   1128

CCA ATG GTA AAT GGA AGG GCC ATT ATA AAT TTC CAA CAC TTG TCA ACA TAT GAG CCA GAA CAG TTT GAG GTT   3456
 P   M   V   N   G   R   A   I   I   N   F   Q   H   L   S   T   Y   E   P   E   Q   F   E   V   1152

GCA AAC ACC CGG TCG ACT CAA GAA CAG TTT CAA GCA TGG TAT GAG GGA GTG AAA GGG GAC TAT GGT GTT GAC   3528
 A   N   T   R   S   T   Q   E   Q   F   Q   A   W   Y   E   G   V   K   G   D   Y   G   V   D   1176

GAT ACA GGA ATG GGG ATC TTA TTG AAT GGA TTA ATG GTT TGG TGC ATT GAA AAT GGC ACA TCC CCA AAT ATA   3600
 D   T   G   M   G   I   L   L   N   G   L   M   V   W   C   I   E   N   G   T   S   P   N   I   1200

AAT GGC GTG TGG ACT ATG ATG GAT GGT GAT GAG CAA GTG ACA TAT CCA ATT AAA CCA TTG TTG GAC CAT GCA   3672
 N   G   V   W   T   M   M   D   G   D   E   Q   V   T   Y   P   I   K   P   L   L   D   H   A   1224

GTG CCT ACT TTT AGG CAG ATT ATG ACG CAC TTC AGT GAC GTT GCT GAA GCT TAC ATA GAA ATG CGA AAC CGT   3744
 V   P   T   F   R   Q   I   M   T   H   F   S   D   V   A   E   A   Y   I   E   M   R   N   R   1248

ACA AAG GCG TAC ATG CCG AGG TAT GGT CTA CAA CGT AAT TTG ACT GAT ATG AGT CTT GCG CGA TAT GCA TTT   3816
 T   K   A   Y   M   P   R   Y   G   L   Q   R   N   L   T   D   M   S   L   A   R   Y   A   F   1272

GAT TTC TAC GAG CTG CAT TCA ACC ACC CCT GCA CGT GCA AAA GAA GCA CAT TTA CAG ATG AAG GCA GCC GCG   3888
 D   F   Y   E   L   H   S   T   T   P   A   R   A   K   E   A   H   L   Q   M   K   A   A   A   1296

CTT AAG AAT GCG AAA AAT CGG TTG TTT GGT TTG GAC GGA AAC GTC TCC ACG CAA GAA GAA GAT ACG GAG AGG   3960
 L   K   N   A   K   N   R   L   F   G   L   D   G   N   V   S   T   Q   E   E   D   T   E   R   1320

CAC ACG ACA ACT GAT GTT ACT AGA AAT ATA CAT AAC CTC TTA GGA ATG AGG GGT GTG CAA TAG GAC ATC CTC   4032
 H   T   T   T   D   V   T   R   N   I   H   N   L   L   G   M   R   G   V   Q   *               1341

TGC ACT GTA GTT TAT ACT TAT GTT ATC TTT AGT ATG CCT TTA ATT TAA ATT CGT GTC TTT CAG TCC CGA AGG   4104

AGA TGG TTG AGT GCA TAA CAT GGT GGG ATT ATA TCT CGG TTA TTG CAT TTG AGA AGT CGC CTT TCT ATT ACG   4176

TAT CAT AAG GGA CTC TTA AAA GTG AGG AGT ACC TCG TAA GAA AAG CCT TTT TGG TTC GTG ATC GAG CC -(A)n  4244
```

Fig. 2. Nucleotide sequence of the 3′-terminal region of SPFMV-S. The predicted amino acid sequence of the open reading frame coding for the putative polyprotein is shown. Horizontal arrows below the nucleotide sequence indicate an inverted repeat [15]. The amino acids which form the potential CI/6K_2/NIa-VPg/NIa-Pro/NIb/CP cleavage sites proposed in this article are double underlined and the cleaved peptide bonds are indicated by vertical arrows. The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GenBank Nucleotide Sequence Databases under the accession number D38543

Table 1. The deduced cleavage sites recognized by the NIa proteinase in the SPFMV-S polyprotein

CI/6K$_2$	SCVLHQ/T
6K$_2$/NIa	QTVQHQ/G
NIa/NIb	IQVYAQ/T
NIb/CP	LEVYHQ/S
NIa internal cleavage site	SLVGHE/S
	S
Consensu	V- HQ/T
cleavage site	A G

Sites were identified by comparison with the determined and/or proposed sites in several potyvirus polyproteins [23]

Table 2. Comparison of SPFMV-S proteins with those of other potyviruses

			NIa			
		6K$_2$	VPg	Pro	NIb	CP
TEV	[2]	42%	49%	53%	64%	50%
TVMV	[6]	30%	51%	49%	60%	44%
PVY	[24]	43%	52%	50%	61%	50%
PPV	[14]	51%	57%	65%	68%	54%
PSbMV	[11]	34%	46%	44%	59%	48%
PeMV	[30]	44%	52%	51%	60%	49%
PRSV	[31]	25%	50%	48%	60%	49%
TuMV	[20]	38%	58%	63%	66%	52%
JGMV	[9]	36%	50%	36%	60%	47%
PVA	[21]	38%	52%	56%	60%	51%

strains, respectively. Usugi et al. showed that the reaction of SPFMV-S with antiserum against SPFMV-RC in PAS-ELISA was weaker than that of SPFMV-O [27], indicating that the S strain is serologically more closely related to the O strain than the RC strain. This suggests that the amino acid substitutions at position 13 (N and D) and 21–23 (INP and NIF) in Fig. 3, found in the coat proteins of the S and RC strains, may affect the antigen/antibody interactions.

A DAG amino acid triplet is found close to the N-terminus (Fig. 3), which is conserved in aphid-transmissible potyviruses and has been shown to be involved in the aphid transmissibility [3, 4]. SPFMV-S is reported to be highly transmitted by aphids [29].

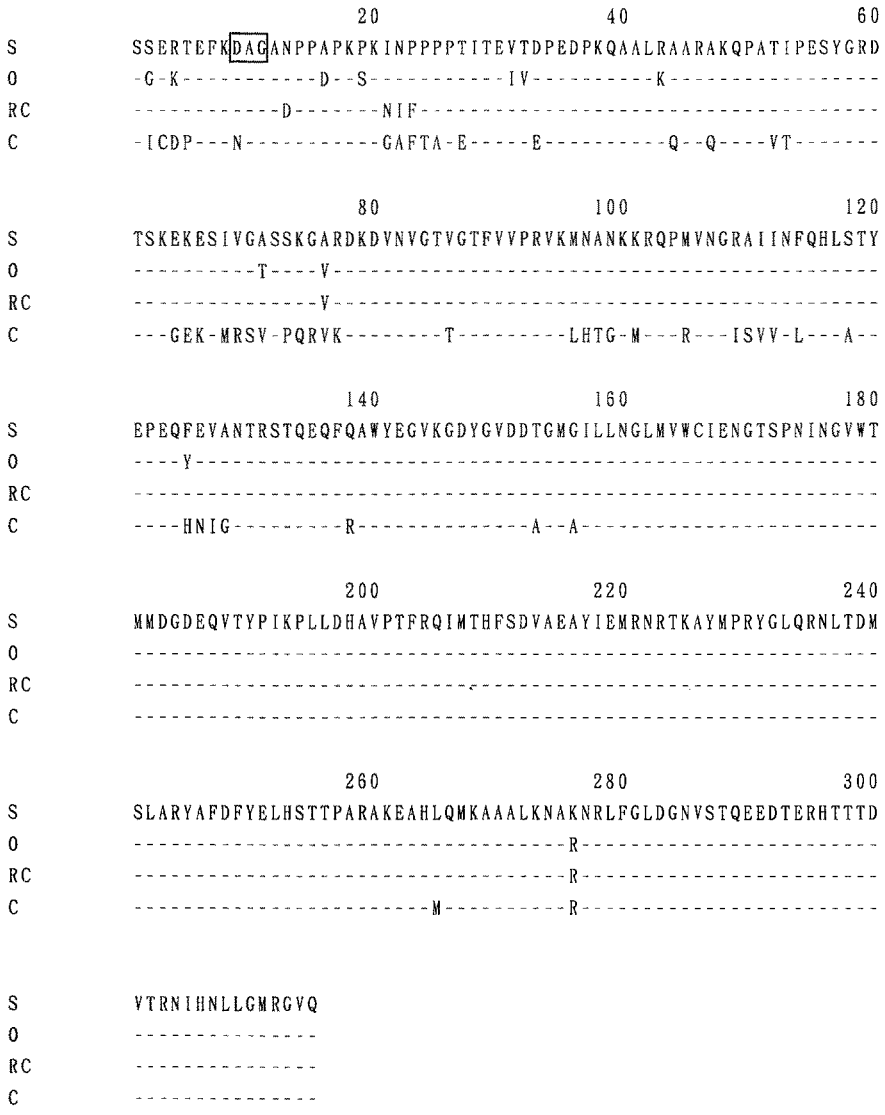

Fig. 3. Amino acid alignment in the region of the coat protein of SPFMV-S, -O [15], -RC and -C [1] strains. Bars indicate identical amino acid residues. DAG triplet [3] is boxed

The NIb protein of SPFMV-S shows 59 to 68% identity with those of other potyviruses (Table 2). On the basis of sequence similarity with other positive-stranded viruses, NIb was proposed to be the RNA-dependent RNA polymerase [23]. The consensus motif [S (T) GXXXTXXXNS (T) (18 to 37 a.a.) GDD], which is conserved in a variety of both animal and plant positive-stranded RNA viral RNA-dependent polymerases [12], is present in the deduced NIb protein in a position similar to that of other potyviruses, starting at position 814 (Fig. 2).

The NIa protein is divided into two domains, the N-terminal VPg domain and the C-terminal proteinase domain [23]. The internal cleavage site is also found in the NIa protein of SPFMV-S, yielding a N-terminal VPg of 192 amino

acids and a C-terminal protease of 243 amino acids. The former domain shows 46% to 58% identity and the latter 36% to 65% identity with the corresponding domains of other potyviruses (Table 2). The catalytic triad of NIa proteinase, His (H), Asp (D) and Cys (C) [8] is conserved in SPFMV-S (polyprotein positions 307, 342 and 412; Fig. 2). Alignment of the potyviral polyprotein sequences shows that the location and spacing of these three amino acids is strictly conserved in the potyviral NIa proteases including that of SPFMV-S. A Tyr (Y) residue in the VPg domain, which links the TVMV VPg to the genomic RNA [19], is also conserved in the potyviral VPg domains including that of SPFMV-S (polyprotein position 132; Fig. 2).

The $6K_2$ protein shows 25% to 51% identity with those of other potyviruses (Table 2). The function of the $6K_2$ protein is still uncertain. Recently, it was reported that the $6K_2$ protein of the TEV is membrane associated and involved in viral replication [22]. A hydrophobicity profile [13] of the $6K_2$ protein of SPFMV-S showed a central hydrophobic domain (data not shown), which is structurally conserved in all potyvirus $6K_2$ proteins and may function as an anchor by direct insertion into the lipid bilayer [22].

Acknowledgements

We thank Junko Ikegami for her technical assistance. This work was supported by Grants-in-Aid "Integrated Research Program for the Use of Biotechnological Procedures for Plant Breeding" and "Research Project of Promoting Biotechnology for Prefectures" from the Ministry of Agriculture, Forestry and Fisheries of Japan.

References

1. Abad JA, Conkling MA, Moyer JW (1992) Comparison of the capsid protein cistron from serologically distinct strains of sweet potato feathery mottle virus (SPFMV). Arch Virol 126: 147–157
2. Allison R, Johnston RE, Dougherty WG (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. Virology 154: 9–20
3. Atreya CD, Raccah B, Pirone TP (1990) A point mutation in the coat protein abolishes aphid transmissibility of a potyvirus. Virology 178: 161–165
4. Atreya PL, Atreya CD, Pirone TP (1991) Amino acid substitutions in the coat protein result in loss of insect transmissibility of a plant virus. Proc Natl Acad Sci USA 88: 7887–7891
5. Beachy RN, Loesch-Fries S, Tumer NE (1990) Coat protein-mediated resistance against virus infection. Ann Phytopathol 28: 451–474
6. Domier LL, FranklinI KM, Shahabuddin M, Hellmann GM, Overmeyer H, Hiremath ST, Siaw MFE, Lomonossoff GP, Shaw JG, Rhoads RE (1986) The nucleotide sequence of tobacco vein mottling virus RNA. Nucleic Acids Res 14: 5417–5430
7. Cali BB, Moyer JW (1981) Purification, serology, and particle morphology of two russet crack strains of sweet potato feathery mottle virus. Phytopathology 71: 302–305

8. Dougherty WG, Parks TD, Caru SM, Bazan JF, Fletterich RJ (1989) Characterization of the catalytic residues of the tobacco etch virus 49 kDa proteinase. Virology 172: 302–310
9. Gough KH, Shukla DD (1993) Nucleotide sequence of Johnsongrass mosaic potyvirus genomic RNA. Intervirology 36: 181–192
10. Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. Gene 25: 263–269
11. Johansen E, Rasmussen OF, Heide M, Borkhardt B (1991) The complete nucleotide sequence of pea seed-borne mosaic virus RNA. J Gen Virol 72: 2625–2632
12. Kamer G, Argos P (1984) Primary structural comparison of RNA-dependent RNA polymerases from plant, animal and bacterial viruses. Nucleic Acids Res 12: 7269–7282
13. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132
14. Maiss E, Timpe U, Brisske A, Jelkmann W, Casper R, Himmler G, Mattanovich D, Katinger HWD (1989) The complete nucleotide sequence of plum pox virus RNA. J Gen Virol 70: 513–524
15. Mori M, Usugi T, Hayashi T, Nishiguchi M (1994) Nucleotide sequence at the 3′-terminal region of sweet potato feathery mottle virus (ordinary strain, SPFMV-O) RNA. Biosci Biotech Biochem 58: 965–967
16. Moyer JW, Kennedy GG (1978) Purification and properties of sweet potato feathery mottle virus. Phytopathology 68: 998–1004
17. Moyer JW, Cali BB (1985) Properties of sweet potato feathery mottle virus RNA and capsid protein. J Gen Virol 66: 1185–1189
18. Moyer JW, Salazar LF (1989) Virus and viruslike diseases of sweet potato. Plant Dis 73: 451–455
19. Murphy JF, Rychlik W, Rhoads RE, Hunt AG, Shaw JG (1991) A tyrosine residue in the small nuclear inclusion protein of tobacco vein mottling virus links the VPg to the viral RNA. J Virol 65: 511–513
20. Nicolas O, Laliberte J-F (1992) The complete nucleotide sequence of turnip mosaic potyvirus RNA. J Gen Virol 73: 2785–2793
21. Puurand U, Makinen K, Paulin L, Saarma M (1994) The nucleotide sequence of potato virus A genomic RNA and its sequence similarities with other potyviruses. J Gen Virol 75: 457–461
22. Restrepo-Hartwig MA, Carrington JC (1994) The tobacco etch potyvirus 6-kilodalton protein is membrane associated and involved in viral replication. J Virol 68: 2388–2397
23. Riechmann JL, Lain S, Garcia JA (1992) Highlights and prospects of potyvirus molecular biology. J Gen Virol 73: 1–16
24. Robaglia C, Durand-Tardif M, Tronchet M, Boudazin G, Astier-Manifacier S, Casse-Delbart F (1989) Nucleotide sequence of potato virus Y (N strain) genomic RNA. J Gen Virol 70: 935–947
25. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467
26. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
27. Usugi T, Onuki M, Hayashi T (1990) Detection of some viruses by protein A sandwich ELISA (PAS-ELIZA). Kyushu Agric Res 53: 106 [Abstract in Japanese]
28. Usugi T, Nakano M, Shinkai A, Hayashi T (1991) Three filamentous viruses isolated from sweet potato. Ann Phytopathol Soc Japan 57: 512–521
29. Usugi T, Nakano M, Onuki M, Maoka T, Hayashi T (1994) A new strain of sweet potato feathery mottle virus that causes russet crack on fleshy roots of some Japanese cultivars of sweet potato. Ann Phytopathol Soc Japan 60: 545–554

30. Vance VB, Moore D, Turpen TH, Bracker A, Hollowell VC (1992) The complete nucleotide sequence of pepper mottle virus genomic RNA: comparison of the encoded polyprotein with those of other sequenced potyviruses. Virology 191: 19–30
31. Yeh S-D, Jan F-J, Chiang C-H, Doong T-J, Chen M-C, Chung P-H, Bau H-J (1992) Complete nucleotide sequence and genetic organization of papaya ringspot virus RNA. J Gen Virol 73: 2531–2541

Authors' address: Dr. M. Mori, Department of Molecular Biology, National Institute of Agrobiological Resources, 2-1-2, Kan-nondai, Tsukuba, Ibaraki 305, Japan.

Received December 5, 1994