Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Structure of the Glycoprotein Gene of Sonchus Yellow Net Virus, a Plant Rhabdovirus[1]

KAREN-BETH GOLDBERG, BRETT MODRELL,[2] BRADLEY I. HILLMAN,[3]
LOUIS A. HEATON,[4] TAE-JIN CHOI, AND A. O. JACKSON[5]

Department of Plant Pathology, University of California, Berkeley, California 94720

The nucleotide sequence of the glycoprotein (G) gene of sonchus yellow net virus (SYNV), a plant rhabdovirus, was determined from viral genomic and mRNA cDNA clones. The G cistron is 2045 nucleotides (nt) long and the G protein mRNA open reading frame (ORF), as determined from the cDNA sequence, contains 1896 nt and encodes a protein of 632 amino acids. Immunoblots with antibodies elicited against the purified glycoprotein from virus particles reacted with a fusion protein produced in *Escherichia coli*, indicating that the cloned ORF encodes the G protein. The 5' end of the G protein mRNA corresponds to nt 5111, relative to the 3' end of the viral (minus sense) genome, as determined by primer extension from mRNA isolated from infected plants, and extends to nt position 7155 on the genomic RNA. A 34-nt untranslated 5' leader sequence and a 115-nt untranslated 3' end flank the ORF on the mRNA. The gene junctions on either side of the G gene on the genomic RNA are identical to those previously described for other SYNV genes and are similar to sequences separating genes of animal rhabdoviruses. The predicted molecular weight of the G protein is 70,215 Da, a value less than the 77,000 Da estimated for the glycosylated G protein from virus particles. The deduced amino acid sequence of the SYNV G protein shares little direct relatedness with the G proteins of other rhabdoviruses, but appears to contain a similar signal sequence, a transmembrane anchor domain, and glycosylation signals. In addition, the SYNV G protein contains a putative nuclear targeting site near the carboxy terminus, which may be involved in transit to the nuclear membrane prior to morphogenesis. © 1991 Academic Press, Inc.

INTRODUCTION

Sonchus yellow net virus (SYNV), a plant rhabdovirus, has bacilliform particles (Jackson and Christie, 1977, 1979) which consist of a distinctive nucleocapsid core and an envelope (Selstam and Jackson, 1983). Physicochemical studies of the virus and analysis of RNAs from infected plants reveal that the genome consists of a minus sense RNA approximately 13,000 nucleotides (nt) long (Milner and Jackson, 1979; Rezaian et al., 1983). The four major proteins observed in purified preparations of SYNV are the nucleocapsid (N), which is tightly associated with the genomic RNA, and a putative phosphoprotein (M2), which possibly associates with the nucleocapsid protein, plus a matrix (M1) protein, which in conjunction with the glycoprotein (G) is thought to form the viral envelope (Jackson, 1978).

We have previously reported the characterization and sequences of the N (Zuidema et al., 1987), M1 (Hillman et al., 1990), and M2 (Heaton et al., 1987) protein genes. Two other incompletely characterized genes most likely represent the replicase (L) protein (Choi et al., in preparation) and a unique protein (sc-4) that has not been detected in virion preparations and is probably a nonstructural protein (Goldberg, unpublished results).

Mapping and sequencing of cDNA clones of the SYNV genomic and messenger RNAs has revealed that the gene order is similar to that of animal rhabdoviruses (Heaton et al., 1989). A 144-nt leader RNA(l) is transcribed from the 3' terminus of the negative strand genomic RNA and is followed by six open reading frames that are expressed by transcription of positive sense polyadenylated mRNAs (Milner and Jackson, 1979; Heaton et al., 1989). The gene order from the 3' to 5' ends of the genomic RNA is l-N-M2-sc4-M1-G-L (Fig. 1) and each gene is separated by a 16-nucleotide consensus gene junction sequence (Heaton et al., 1989) which contains the transcriptional start and stop sites. The SYNV G protein is associated with the viral envelope (Jackson, 1978) and is thought to interact with the matrix protein, M1, during morphogenesis. The G protein is glycosylated (Jackson, 1978) as revealed by periodic acid-Schiff staining. Recent experiments demonstrate that the nonglycosylated deriva-
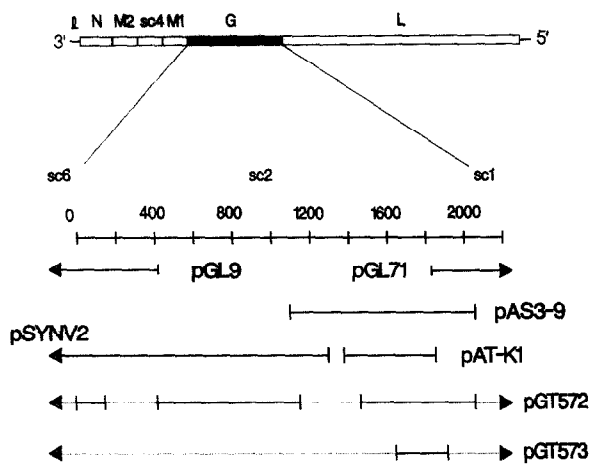
FIG. 1. Strategy for sequencing the SYNV complementary (sc) RNA 2 gene. The map at the top of the figure represents the organization of the SYNV negative strand genome in the 3' to 5' orientation. The headings I, N, M2, sc4, M1, G, and L refer to the 144-nt leader transcript, the nucleocapsid protein gene, the M2 protein gene, the scRNA-4 gene thought to encode a nonstructural protein, the M1 (scRNA-6) protein gene, the G protein gene, and the L (scRNA-1) protein gene, respectively. The sequenced region (scRNA-2) corresponding to the G protein gene begins at position 5111 and extends to position 7155 from the 3' end of the genomic RNA. The recombinant RNA plasmids pGL9, pGL71, and pSYNV2 were derived from SYNV genomic RNA and pAT-K1 and pAS3-9 were derived from poly(A)$^+$ from SYNV-infected plants. The plasmids are positioned on the map according to their genomic locations. All clones were completely sequenced on both strands by primer-based dideoxynucleotide sequencing. In addition, two other G protein gene clones, pGT572 and pGT573, spanning the entire cistron, were partially sequenced (as indicated by the solid line). The lines with arrows indicate that the clones bridged a junction between the G gene sequence and an adjacent SYNV gene.

tive synthesized in tunacamycin-treated protoplasts migrates in polyacrylamide gels slightly more rapidly than the glycosylated form (Jones and Jackson, 1990). Moreover, the glycosylated derivative has a central role in morphogenesis of SYNV, since normal accumulation of enveloped bacilliform particles in perinuclear spaces surrounding the nucleus (Christie et al., 1974) is interrupted in tunacamycin-treated protoplasts, where nonenveloped nucleoprotein cores accumulate in arrays around the inner periphery of the nucleus (van Beek et al., 1985). These results are compatible with a model for morphogenesis of many plant rhabdoviruses, including SYNV, in which the nucleocapsid assembles in the nucleus and buds through the inner nuclear membrane (Francki, 1973; Jackson et al., 1987). In contrast, animal rhabdoviruses such as vesicular stomatitis virus (Wagner, 1987) and rabies virus (Wunner, 1987) bud through the plasma membrane or cytoplasmic membrane, respectively.

To gain a more detailed understanding of the structure and function of the SYNV G protein, we have se-

quenced the G protein gene. The deduced amino acid sequence indicates that the G protein contains domains similar to the signal sequence, glycosylation sites, and transmembrane anchor domains of animal rhabdovirus G proteins. A possible nuclear targeting signal has also been identified near the carboxy terminus of the G protein.

MATERIALS AND METHODS

General

The methods used for SYNV propagation, maintenance, purification, and RNA extraction were previously described (Jackson and Christie, 1977; Milner and Jackson, 1979).

Construction of SYNV-specific cDNA clones and characterization of the 5' terminus of the mRNA

Five clones obtained from SYNV genomic or poly(A)$^+$ RNA hybridized to the SYNV complementary RNA-2 (scRNA-2), the putative G protein mRNA (Heaton et al., 1989). These clones were designated pGL9, pSYNV2, pAS3-9, pGL71, and pAT-K1 (Fig. 1). In addition, two G protein clones (pGT572 and pGT573), constructed 2 years after the initial clones, were partially sequenced to provide confirmatory sequence data. Sequencing was performed with Sequenase (USB Corp., Cleveland, OH), using the dideoxynucleotide method on both strands of these plasmids by using sequence-specific primers on either single-stranded or double-stranded DNA according to the protocol suggested by the supplier. Oligo-nucleotide primers were synthesized using an Applied Biosystems automated DNA synthesizer by the University of California—Berkeley Department of Biochemistry. To identify the transcriptional start site for the mRNA, a 20-nucleotide oligomer complementary to the sc-2 mRNA was end-labeled with ^{32}P and used to prime reverse transcription of poly(A)$^+$ mRNA from SYNV-infected plants (Zuidema et al., 1987; Heaton et al., 1987; Hillman et al., 1990).

Preparation of G-specific antibodies and identification of the G protein

Viral proteins of SYNV were dissociated with 2% SDS and separated on SDS–polyacrylamide gels (Laemmli, 1970). The gels were stained with Coomassie blue, and the G protein band was excised and electroeluted overnight from polyacrylamide gels. This protein was further processed through an Amicon Centricon 30 column by using 3 ml of phosphate buffered saline, pH 7.4, during each of three centrifugation cycles (5000 rpm for 20 min in a Sorvall SS-34 rotor). A 1:1 (v/v) emulsion of Ribi adjuvant in 0.9% NaCl (Ribi

ImmunoChem Research, Inc., Hamilton, MT) and the eluted G protein were injected into four female Swiss Webster mice (15 μg SYNV G protein/mouse) at Days 1, 7, and 21. Development of ascites fluid and recovery of antibodies were carried out essentially as described by Hunter *et al.* (1990).

In order to verify that the sequenced gene corresponded to the G protein gene, fusion proteins were constructed by ligating a portion of the sequenced gene to the carboxy terminus of the β-galactosidase gene. For this purpose, a fragment at the 3′ end of the putative G protein gene was inserted into the *Bam*HI site of the pUR290 plasmid series (Ruther and Muller-Hill, 1983) to yield a fusion product containing 234 amino acids from the carboxy-terminus of the G protein. Expression of the chimeric protein, polyacrylamide gel electrophoresis, and immunoblotting have been described previously (Hunter *et al.*, 1990; Zuidema *et al.*, 1987).

Computer analyses of nucleotide and amino acid sequences

The nucleotide sequence data were analyzed for translational open reading frames and restriction endonuclease sites. For amino acid sequence analyses essentially the same parameters described by Heaton *et al.* (1987) were used for the Beckman Microgenie program and the NIH-sponsored "Bionet" National Computer Resource for Molecular Biology (Grant 1 U41 RR-01685-03).

RESULTS AND DISCUSSION

Identification of the 5′ terminus of scRNA-2

The 5′ end of SYNV mRNA sc-2 (Rezaian *et al.*, 1983), which we have postulated to correspond to the G gene mRNA (Heaton *et al.*, 1989), was determined by extension of a ^{32}P end-labeled DNA primer that was complementary to the 5′ proximal end of the mRNA. Extension of this primer revealed a doublet corresponding to the 5′ end of the SYNV mRNA (Fig. 2). Similar results were obtained with primer extensions of the nucleocapsid (N) protein mRNA of SYNV (Zuidema *et al.*, 1987), the M2 protein mRNA (Heaton *et al.*, 1987), and the M1 protein mRNA (Hillman *et al.*, 1990), suggesting that the first three residues of the scRNA-2 transcripts, beginning at nt position 5111 on the SYNV genomic RNA, are 5′-AAC-3′ as has also been reported for the other SYNV mRNAs (Heaton *et al.*, 1989). There is no evidence that the spacer dinucleotide (GG) on the genomic RNA, which is a component of the gene-junction region at nt positions 5109–5110, is transcribed.
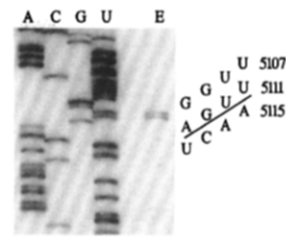

FIG. 2. Identification of the 5′ terminus of the SYNV glycoprotein (sc-2) mRNA. A synthetic ^{32}P end-labeled oligonucleotide (5′-AGTCGTCCTGATATCCTAGC-3′) was extended on poly(A)$^+$ mRNA isolated from SYNV-infected *Nicotiana edwardsonii* plants and the reaction products were analyzed on 6% polyacrylamide gels. The lane adjacent to the primer extension products (E) contains sequence reactions from pSYNV2 primed with the same oligonucleotide. The nucleotide sequence is represented in the minus sense. The first three nucleotides of the underlined region correspond to the consensus rhabdovirus mRNA initiation trinucleotide.

This finding is in agreement with previous sequencing results (Heaton *et al.*, 1989).

Identification of the protein encoded by the scRNA-2

A fusion protein was constructed to provide more direct evidence that the G protein found in virus particles was encoded by scRNA-2. Ascites fluid from mice immunized with the G protein was tested against a β-galactosidase fusion product containing 234 amino acids derived from the scRNA-2 gene. Western blot analyses revealed a positive reaction of the SYNV G protein antibodies to the fusion protein as well as to the G protein from virion preparations (Fig. 3). No cross-reactions occurred with other viral or bacterial proteins. These results demonstrate that scRNA-2 is the mRNA for the SYNV G protein.

The complete nucleotide sequence of the G protein gene

Five overlapping plasmids (Fig. 1) were used to obtain the complete sequence of the gene. These plasmids contain an SYNV-specific sequence derived from genomic RNA or poly(A)$^+$ RNA from infected plants that hybridized to the G gene. The genomic clones pGL9 and pSYNV2 hybridized to the M1 (scRNA-6) and the G genes, pAS3-9 and pAT-K1 hybridized to the G gene and were derived from poly(A)$^+$ libraries, and the genomic clone pGL71 overlaps the G and L (scRNA-1) genes.

Evaluation of the nucleotide sequence of the SYNV G protein gene (Fig. 4) reveals the presence of a 2045-nt-long cistron extending from positions 5111 to 7155, relative to the 3′ end of the viral (minus sense) genome. The 5′ end of the mRNA consists of a tetranucleotide (5′-AACU-3′) which is identical to the tetranucleotide
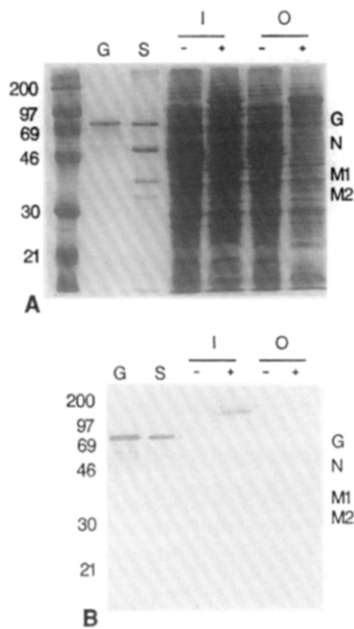
FIG. 3. Analysis of SYNV G protein on sodium dodecyl sulfate (SDS)–polyacrylamide gels. Panels A and B are duplicate gels representing the SYNV glycoprotein (G), the SYNV structural proteins (S), and in-frame (I) or out-of-frame (O) β-galactosidase fusions of a cDNA clone of the G gene with (+) and without (−) induction by IPTG. Protein molecular weight standards are indicated in kilodaltons (Bio-Rad Laboratories, Richmond, CA). Panel A represents a gel stained for total protein with Coomassie brilliant blue. Panel B represents an electrophoretic transfer to nitrocellulose membrane followed by immunoblot analysis with polyclonal ascites fluid containing antibodies to the SYNV G protein. All the proteins were disrupted in 2% SDS and resolved in 12.5% polyacryamide gels.

found at the 5' end of the M1 mRNA (scRNA-6) (Heaton et al., 1989; Hillman et al., 1990). This sequence differs slightly from the 5' ends of the N gene (scRNA-3) and the M2 gene (scRNA-5) which begin with the tetranucleotide 5'-AACA-3' (Zuidema et al., 1987; Heaton et al., 1987). Sequence analysis indicates that an untranslated leader sequence of 34 nt precedes the first AUG (start codon) of the G protein ORF (Fig. 4). As with the other SYNV leader sequences, the G mRNA leader is rich in adenine and uracil, which comprise 27 of the 34 residues. In addition to the translational start AUG at positions 35 to 37, an AUG codon is in the same frame at positions 47 to 49. Neither codon occurs in a more favorable context according to Kozak's rules for translational initiation (Kozak, 1981, 1984), but the first AUG is in a slightly better context when compared to the AUG consensus flanking sequence for plant genes (5'-UAAACAAUGGCU-3') compiled by Joshi (1987) and Lutcke et al. (1987). Usage of the 5'-most AUG would permit synthesis of a putative signal peptide similar in size and hydrophobicity to the proposed glycoprotein signal peptides of VSV, infectious hematopoietic ne-

crosis virus (IHNV), rabies, and Sigma virus of Drosophila (Rose and Gallione, 1981; Koener et al., 1987; Anilionis et al., 1981; Teninges and Bras-Herreng, 1987), so we presume that translation is initiated at this start site.

The translational reading frame, starting with the first AUG, can encode a protein of 632 amino acids from an ORF of 1896 nt (Fig. 4). The first translational stop codon (UGA) is located on the genome at nt position 7041 to 7043 and is followed by two in-phase stop codons (UAA) at nt positions 7068 to 7070 and 7125 to 7127. Alternative reading frames present in the ORF could potentially encode polypeptides no larger than 47 amino acids.

The 3' end of the mRNA, excluding the poly(A) tract, corresponds to position 7155 on the SYNV genome as determined by evaluation of a cDNA clone (pAS3-9), derived by oligo-dT priming of poly(A)$^+$ RNA from SYNV-infected Nicotiana edwardsonii plants. The pAS3-9 clone contains the consensus mRNA gene junction 3' end (5'-AUAAGAAAAA-3') plus a polyadenylic stretch of 26 nt. Thus, the G protein mRNA, excluding the poly(A) tail, is 2045 nt long and represents the second largest cistron on the SYNV genome.

As is the case for the N protein mRNA (scRNA-3) (Zuidema et al., 1987), the G protein mRNA (scRNA-2) has a putative polyadenylation signal (AAUAAA) beginning 33 nt upstream from the poly(A) stretch at nucleotide position 7123. Since SYNV mRNAs are presumably transcribed in the nucleus (Jackson et al., 1987) and the N protein mRNA has a polyadenylation signal (Zuidema et al., 1987), we previously postulated that SYNV might utilize this signal to initiate polyadenylation by host enzymes. However, since the M2 (Heaton et al., 1987) and M1 (Hillman et al., 1990) protein mRNAs lack a polyadenylation signal, the occurrence of this signal may be fortuitous. The animal rhabdoviruses that have been analyzed (Rose and Gallione, 1981; Tordo et al., 1986; Koener et al., 1987; Teninges and Bras-Herreng, 1987; Wagner, 1987) have a cytoplasmic mode of replication and lack the putative polyadenylation signal, yet the mRNAs of these viruses are polyadenylated. Moreover, evidence has been accumulated (Banerjee, 1987; Emerson, 1987) suggesting that polyadenylation of VSV occurs via polymerase slippage or "chattering" along the uracil tract of the gene-junction sequence. We have no direct evidence for such a mechanism with SYNV, but the similarity of the SYNV gene-junction sequences to those of other rhabdoviruses (Heaton et al., 1989) suggests that the members of this group all utilize common mechanisms for termination of transcription, polyadenylation, and initiation of downstream mRNA transcription.

In addition to the primary clones, two recently iso-

```
5111 AACTTTGATAGTTTATTATTAAGAACTTTTCGTAATGTCTCATATAATGAACCTTTTGGTG
                       M  S  H  I (M) N  L  L  V

5172 ATTTCTTTTGTTTTAGCAGGTTCATCATGGTCTTTGCTAGGATATCAGGACGACTTCTCA
  10 I  S  F  V  L  A  G  S//S  W  S  L  L  G  Y  Q  D  D  F  S

5232 TCGAAAAGGTCCGGAGCTTTAGCATCAAACCCGACTTACAATCTCCCCCAAGACAAAGGA
  30 S  K  R  S  G  A  L  A  S  N  P  T  Y  N  L  P  Q  D  K  G

5292 TACGGAAGAGATATGTACCAGCCTTACTACATATGCGAACCAGATAATGATGGGAGTGCA
  50 Y  G  R  D  M  Y  Q  P  Y  Y  I  C  E  P  D  N  D  G  S  A

5352 CTGACTTTGCCATCATGGCATTATTCCTGTAAGGAGTCTTGTATGGGAAACCATCTCAAG
  70 L  T  L  P  S  W  H  Y  S  C  K  E  S  C  M  G  N  H  L  K

5412 AGAGTAGTCAATATAACTGGGGCAAGATGGAATTATGTGGGAATAAGTATACCAGTATTC
  90 R  V  V  N  I  T  G  A  R  W  N  Y  V  G  I  S  I  P  V  F

5472 AAGATAGTGACAAACGAAGTATGCTATACATCCCATGAAAATGTATGGGGTTACTGCTCC
 110 K  I  V  T  N  E  V  C  Y  T  S  H  E  N  V  W  G  Y  C  S

5532 CAATATCAGATTAGCCGTCCGGTTGCAACCCAGAAGAGCGATGTGTCCTGCATAACCTCG
 130 Q  Y  Q  I  S  R  P  V  A  T  Q  K  S  D  V  S  C  I  T  S

5592 TCAATGTGGGACAATGATAAATCACCCATCGGTAGTCTTTATAATATAGTGAACTCAAAT
 150 S  M  W  D  N  D  K  S  P  I  G  S  L  Y  N  I  V  N  S  N

5652 GAGGCAGAGTGTGACTACTTCTCCGACATAACTGACTGCAACAGGGATTATCAGATATTC
 170 E  A  E  C  D  Y  F  S  D  I  T  D  C  N  R  D  Y  Q  I  F

5712 AAGAGAGAGGGGAAGCTAATAAAGAGGAGTGATGATTCTCCATTGGAGCTCTCTATTGTA
 190 K  R  E  G  K  L  I  K  R  S  D  D  S  P  L  E  L  S  I  V

5772 ACGGATGGAATCAGGACAGATCCGGCCTCGGAGTATTTGAGTCTAGATGATGTATCTTGG
 210 T  D  G  I  R  T  D  P  A  S  E  Y  L  S  L  D  D  V  S  W

5832 TTTTGGAAGTTGCCGAACAACGACATGTCGCCCCCATGTGGGTGGGAAAAGACTCAGAAA
 230 F  W  K  L  P  N  N  D  M  S  P  P  C  G  W  E  K  T  Q  K

5892 TTATCTTGTTCTTATACCGATACGACAGATGTTATCAAGTGTAATAGTATAGGCTATACA
 250 L  S  C  S  Y  T  D  T  T  D  V  I  K  C  N  S  I  G  Y  T

5952 TACAACATTCAAGGAATATCTAAGAAGAGCACTTGTGCAGGCAATATATATGACACAGAC
 270 Y  N  I  Q  G  I  S  K  K  S  T  C  A  G  N  I  Y  D  T  D

6012 GGTCCATTTCCTTTCTTCTATGATGCAGAAGAAGCTCTGATGTCCACTGATGATGCTTGC
 290 G  P  F  P  F  F  Y  D  A  E  E  A  L  M  S  T  D  D  A  C

6072 GGAAAAGCGAAACAGGGGAAGCCAGACGCAGACATCGCGTTTATAGAAGGGGTCAACAGG
 310 G  K  A  K  Q  G  K  P  D  A  D  I  A  F  I  E  G  V  N  R

6132 GCATTTGAAGATCTAGAGTTGACTTATTGTTCTGCTACATGTGACCTCTTTGCAAGGCAA
 330 A  F  E  D  L  E  L  T  Y  C  S  A  T  C  D  L  F  A  R  Q
```

```
6192 GGAACTCCTAATGAAGATCATGTCCTAGATACCCCGATAGGAACATGGAGATATGTTATG
 350 G  T  P  N  E  D  H  V  L  D  T  P  I  G  T  W  R  Y  V  M

6252 AGAGATAATCTGGATCCAGCCCTCGTTCCTTGTCTTCCTACATCAAACTGGACGATCAGC
 370 R  D  N  L  D  P  A  L  V  P  C  L  P  T  S  N  W  T  I  S

6312 GATCCCACCACCATATGTCATGGGAAAGACCACATACTCGTAGTTGACACAGCAACTGGA
 390 D  P  T  T  I  C  H  G  K  D  H  I  L  V  V  D  T  A  T  G

6372 CACTCAGGAAGTTGGGATACAAAGAAAGACTATATAATAACAGGGGAAGTGTGCAACACT
 410 H  S  G  S  W  D  T  K  K  D  Y  I  I  T  G  E  V  C  N  T

6432 AATAACGACGAGATGGGAGATGATTATGACGGGATGCGGGACAAAATATTAAGAGGAGAA
 430 N  N  D  E  M  G  D  D  Y  D  G  M  R  D  K  I  L  R  G  E

6492 ACCATAGAAATCAAATTCTGGACAGGCGATATAATCAGGATGGCACCACCCTACGACAAC
 450 T  I  E  I  K  F  W  T  G  D  I  I  R  M  A  P  P  Y  D  N

6552 CCAGAATGGATCAAAGGTTCAGTTCTATTCCGTCAGAACCCCGGCTGGTTCTCCAGTGTA
 470 P  E  W  I  K  G  S  V  L  F  R  Q  N  P  G  W  F  S  S  V

6612 GAATTGAACAAGGACATGATCCATACCAGAGATAACATAACAGACTTGCTCACCGTGATG
 490 E  L  N  K  D  M  I  H  T  R  D  N  I  T  D  L  L  T  V  M

6672 GTTCAAAATGCAACGGCAGAGGTCATGTATAAGCGATTAGATCCTAAGACCATGAAGCAC
 510 V  Q  N  A  T  A  E  V  M  Y  K  R  L  D  P  K  T  M  K  H

6732 ATCCTGTTCGCCGAGATTGTAGATGGAGTAGGCAACGTGTCAGGTAAGATCAGCGGATTT
 530 I  L  F  A  E  I  V  D  G  V  G  N  V  S  G  K  I  S  G  F

6792 CTAACTGGTCTATTCGGAGGGTTCACAAAAGCTGTGATAATAGTGGCATCCCTTGCTATA
 550 L  T  G  L  F  G  G  F  T  K  A  V  I  I  V  A  S  L  A  I

6852 TGCTACATTGTGCTAAGTGTACTATGGAAGGTGAGGCTAGTTGCATCCATATTCAACAGC
 570 C  Y  I  V  L  S  V  L  W  K  V  R  L  V  A  S  I  F  N  S

6912 GCCAAGAAAAAGAGAGTCAGGATATCCGATATCCTAGATGAAGAACCCCATAGGATCCAG
 590 A  K  K  K  R  V  R  I  S  D  I  L  D  E  E  P  H  R  I  Q

6972 CAGAGTCGACCAACTCTGTCCAGGAAGAAGAAAACCAGAGAGAGCATTCAGATGCTTCTG
 610 Q  S  R  P  T  L  S  R  K  K  K  T  R  E  S  I  Q  M  L  L

7032 AACGACATCTGAATCCACCCCATAAACACGACCAACTAACTAGCCCTCCTAGTCATATTC
 630 N  D  I  *

7092 AACATCGATCAAACGACCCATCCTAATATGAAATAAAATTACGGCCACTATCATATAAGA

7152 AAAA
```

FIG. 4. Nucleotide and deduced amino acid sequences of the SYNV G protein gene numbered relative to the 3' terminus of the minus sense SYNV genome. The nucleotide sequence is presented in the positive sense and corresponds to positions 5111 to 7155. The slashed area (//) represents the putative signal sequence cleavage site; the underlined regions correspond to putative glycosylation signals (Asn–X–Ser/Thr), the putative transmembrane anchor domain is in bold type (amino acids 570 to 594), and the bold-underlined region corresponds to a possible nuclear targeting signal. A second possible translational start site is represented by (M), and the asterisk (*) represents the translational stop codon.

lated subclones (pGT572 and pGT573) spanning the entire G gene (Fig. 1) were partially sequenced on both strands to provide confirmatory sequence. The clone pGT572 was sequenced from nt 5111 to 5201, 5522 to 6215, and 6555 to 7158 and differences in the nucleotide sequence were located at nt 5798 (C to A) and at nt 5808 (T to C), which respectively maintained the amino acids alanine and leucine. A third substitution at nt 6046 (C to A) resulted in an amino acid change from alanine to aspartic acid. These changes did not affect the predicted open reading frame of the G gene. However, the addition of an adenine on pGT572 at nt 6922 results in a change of the reading frame deduced from the sequence of pAS3-9 and pAT-K1 and shortens the predicted ORF by 30 amino acids. To provide additional sequence in this region, the other recent G gene clone, pGT573, was sequenced from nt 6789 to 7044 and no additions or substitutions of nucleotides occurred that differed from the originally sequenced material. Since three of four clones are in agreement it

seems likely that the additional adenine observed in clone pGT572 either is an inconsistency that occurred during the replication of the viral genomic RNA in the host plant or is an artifact of cDNA cloning.

Analysis of the G protein

The deduced G protein has little direct relatedness with membrane-associated G proteins of other rhabdoviruses, but some general structural similarities are evident. Our sequencing data reveal that the ORF of the G gene can encode a 70,215-Da protein, which represents the nonglycosylated form. This is in agreement with the observation that protoplasts infected with SYNV in the presence of tunacamycin, an inhibitor of glycosylation, synthesize a G protein of ca. 70,000 M_r (Jones and Jackson, 1990). Glycosylation of the protein is presumably responsible for the reduced electrophoretic mobility to yield an estimated 77,000 M_r for the native G protein synthesized during replication in

the plant cell. Hunter et al. (1990) detected specific binding of four lectins (wheat germ agglutinin, concanavalin A, and the agglutinins of Pisum sativum L. and Lycopersicon esculentum Miller) to the SYNV G protein, indicating the presence of the sugars N-acetyl-β-D-glucosamine and mannose. Six potential glycosylation sites (Asn–X–Ser/Thr) are located at nucleotide positions 5259 to 5267 (amino acids 39 to 41 from the first methionine), 5421 to 5429 (aa 93 to 95), 6297 to 6305 (aa 385 to 387), 6645 to 6653 (aa 501 to 503), 6678 to 6686 (aa 512 to 514), and 6765 to 6773 (aa 541 to 543) (Fig. 4). Chemical analyses of the G protein will be required to determine which of these sites are glycosylated during the infection cycle. The only discernable similarities in location for these glycoprotein signals and the Sigma virus of Drosophila (Teninges and Bras-Herreng, 1987) and IHNV (Koener et al., 1987) are their presence at the amino and carboxy-termini. Rabies virus has three glycosylation signals on the G protein sequence; one has a location identical to that of IHNV at the amino terminal portion of the ORF, and the second and third glycosylation signals flank the membrane domain (Koener et al., 1987). VSV has two glycosylation signals on the G protein sequence, one at the midpoint and the second following the membrane domain, toward the carboxy-terminus (Rose et al., 1981; Koener et al., 1987).

The deduced SYNV G protein sequence also contains a putative proteolytic signal sequence. In general, signal peptides are at least 11 amino acids long and are composed primarily of an uncharged hydrophobic stretch (Perlman and Halvorson, 1983). In SYNV, the predicted G protein translation product from the first AUG possesses a hydrophobic domain at the amino terminus and, as indicated by the slashed (//) region in Fig. 4, following the first 17 amino acids, a characteristic peptidase cleavage site (Ala–X–Ser) appears. IHNV (Koener et al., 1987) and the SYNV G protein share a common putative signal peptide cleavage site at the carboxy-terminus of their putative signal peptides.

A potential transmembrane anchor domain, corresponding to amino acids 570 to 594, is also present at the predicted carboxy-terminus of the G protein (Fig. 4). In general, transmembrane proteins contain three domains: internal, membrane spanning, and external. Typically a transmembrane anchor domain includes a 20- to 27-amino-acid stretch of uncharged, primarily hydrophobic residues (Wiley and Skehel, 1990). The SYNV G protein shares this feature with several other viral glycoproteins including those of VSV (Rose and Gallione, 1981), rabies (Tordo et al., 1986), IHNV (Koener et al., 1987), the Sigma virus of Drosophila (Teninges and Bras-Herreng, 1987), and human coronavirus (Raabe et al., 1990).

Plant rhabdoviruses can be subdivided into at least three groups depending on the site of nucleocapsid formation and assembly of virions (Jackson et al., 1987). Two groups appear to replicate in the cytoplasm, maturing either in association with the endoplasmic reticulum (ER) or in membranes associated with viroplasms, and accumulate in the vesicles of the ER or in vacuole-like sacs, respectively (Jackson et al., 1987). However, the third group of plant rhabdoviruses, which includes SYNV, differs markedly from animal rhabdoviruses because these plant rhabdoviruses clearly undergo nucleocapsid assembly in the nucleus and mature virions accumulate in the perinuclear space formed by the inner and outer nuclear membrane. Cytopathological studies of SYNV-infected plant protoplasts strongly suggest that the virus buds from the inner nuclear envelope into the perinuclear space during the envelopment phase (van Beek et al., 1985). This suggests a model whereby the G protein is localized in the inner leaflet of the perinuclear membrane prior to morphogenesis (Christie et al., 1974; van Beek et al., 1985). The available information suggests that the assembled nucleocapsids accumulate at the patches of G protein and that the virions bud through the inner nuclear membrane into the lumen of the endoplasmic reticulum.

Since SYNV undergoes morphogenesis in association with the nucleus, we attempted to identify possible nuclear targeting sequences in the G protein. Although it is not yet clear exactly what mechanism the SYNV G protein uses to penetrate the inner nuclear envelope, one possible nuclear targeting sequence is evident immediately downstream of the transmembrane anchor domain of the SYNV G protein. This signal begins at amino acid position 591 from the first initiation codon and consists of the region Lys–Lys–Lys–Arg. This sequence appears to be characteristic of the nuclear targeting sequences found within the SV40 T antigen (Kalderon et al., 1984), the infected cell proteins ICP4 and ICP8 of HSV type 1 (McGeoch et al., 1986; DeLuca and Schaffer, 1988; Paterson and Everett, 1988; Knipe, 1989), and the polymerase basic protein 1 of influenza virus (Nath and Nayak, 1990). Additional inspection of the sequences of other SYNV protein genes revealed that the nucleocapsid (Zuidema et al., 1987) and the M1 and M2 proteins (Heaton et al., 1987; Hillman et al., 1990) also contain similar regions that may specify nuclear localization (Goldberg, unpublished data). We are currently initiating studies to investigate the roles of these putative transport signals for membrane localization and viral morphogenesis.

ACKNOWLEDGMENTS

REFERENCES

ANILIONIS, A., WUNNER, W. H., and CURTIS, P. J. (1981). Structure of the glycoprotein gene in rabies virus. *Nature (London)* **294**, 275–277.

BANERJEE, A. K. (1987). Transcription and replication of rhabdoviruses. *Microbiol. Rev.* **51**, 66–87.

CHRISTIE, S. R., CHRISTIE, R. G., and EDWARDSON, J. R. (1974). Transmission of a bacilliform virus of sowthistle and *Bidens pilosa*. *Phytopathology* **64**, 840–845.

DELUCA, N. A., and SCHAFFER, P. A. (1988). Physical and functional domains of the herpes simplex virus transcriptional regulatory protein ICP4. *J. Virol.* **62**, 732–743.

EMERSON, S. U. (1987). Transcription of vesicular stomatitis virus. In "The Rhabdoviruses" (R. R. Wagner, Ed.), pp. 245–269. Plenum, New York.

FRANCKI, R. I. B. (1973). Plant rhabdoviruses. *Adv. Virus Res.* **18**, 257–345.

HEATON, L. A., ZUIDEMA, D., and JACKSON, A. O. (1987). Structure of the M2 protein gene of sonchus yellow net virus. *Virology* **161**, 234–241.

HEATON, L. A., HILLMAN, B. I., HUNTER, B. G., ZUIDEMA, D., and JACKSON, A. O. (1989). The physical map of sonchus yellow net virus genome, a plant rhabdovirus with six genes and conserved gene-junction sequences. *Proc. Natl. Acad. Sci. USA* **86**, 8665–8668.

HILLMAN, B. I., HEATON, L. A., HUNTER, B. G., MODRELL, B., and JACKSON, A. O. (1990). Structure of the gene encoding the M1 protein of sonchus yellow net virus. *Virology* **179**, 201–207.

HUNTER, B. G., RICHARDSON, J., DIETZGEN, R., KARU, A., SYLVESTER, E. S., JACKSON, A. O., and MORRIS, T. J. (1990). Purification and characterization of strawberry crinkle virus. *Phytopathology* **80**, 282–287.

JACKSON, A. O. (1978). Partial characterization of the structural proteins of sonchus yellow net virus. *Virology* **87**, 172–181.

JACKSON, A. O., and CHRISTIE, S. R. (1977). Purification and some physicochemical properties of sonchus yellow net virus. *Virology* **77**, 344–355.

JACKSON, A. O., and CHRISTIE, S. R. (1979). Sonchus yellow net virus. In "Descriptions of Plant Viruses," No. 205. Commonw. Mycol. Inst., Assoc. Appl. Bio., Kew, Surrey, England.

JACKSON, A. O., FRANCKI, R. I. B., and ZUIDEMA, D. (1987). Biology structure and replication of plant rhabdoviruses. In "The Rhabdoviruses" (R. R. Wagner, Ed.), pp. 427–508. Plenum, New York.

JONES, R. W., and JACKSON, A. O. (1990). Replication of sonchus yellow net virus in infected protoplasts. *Virology* **179**, 815–820.

JOSHI, C. P. (1987). An inspection of the domain between the putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.* **15**, 6643–6653.

KALDERON, D., ROBERTS, B. L., RICHARDSON, W. D., and SMITH, A. E. (1984). A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499–509.

KNIPE, D. M. (1989). The role of viral and cellular nuclear proteins in herpes simplex virus replication. In "Advances in Virus Research" (K. Maramorosch, F. A. Murphy, and A. J. Shatkin, Eds.), Vol. 37, pp. 85–123. Academic Press, New York.

KOENER, J. F., PASSAVANT, C. W., KURATH, G., and LEONG, J. (1987). Nucleotide sequence of a cDNA clone carrying the glycoprotein gene of infectious hematopoietic necrosis virus, a fish rhabdovirus. *J. Virol.* **61**, 1342–1349.

KOZAK, M. (1981). Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* **9**, 5233–5252.

KOZAK, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12**, 857–872.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**, 680–685.

LUTCKE, H. A., CHOW, K. C., MICKEL, F. S., MOSS, K. A., KERN, H. F., and SCHEELE, G. A. (1987). Selection of AUG initiation codons differs in plants and animals. *EMBO J.* **6**, 43–48.

McGEOCH, D. J., DOLAN, A., DONALD, S., and BRAUER, D. H. K. (1986). Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nucleic Acids Res.* **4**, 1727–1745.

MILNER, J. J., and JACKSON, A. O. (1979). Sequence complementarity of sonchus yellow net virus RNA with RNA isolated from the polysomes of infected tobacco. *Virology* **97**, 90–99.

NATH, S. T., and NAYAK, D. P. (1990). Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). *Mol. Cell. Biol.* **10**, 4139–4145.

PATERSON, T., and EVERETT, R. D. (1988). Mutational dissection of the HSV-1 immediate-early protein Vmw175 involved in transcriptional transactivation and repression. *Virology* **166**, 186–196.

PERLMAN, D., and HALVORSON, I. (1983). A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol.* **167**, 391–409.

RAABE, T., SCHELLE-PRINZ, B., and SIDDELL, S. G. (1990). Nucleotide sequence of the gene encoding the spike glycoprotein of human coronavirus HCV 229E. *J. Gen. Virol.* **71**, 1065–1073.

REZAIAN, M. A., HEATON, L. A., PEDERSON, K., MILNER, J. J., and JACKSON, A. O. (1983). Size and complexity of polyadenylated RNAs induced in tobacco infected with sonchus yellow net virus. *Virology* **131**, 221–229.

ROSE, J. K., and GALLIONE, C. J. (1981). Nucleotide sequences of the mRNAs encoding the vesicular stomatitis virus G and M proteins determined from cDNA clones containing the complete coding regions. *J. Virol.* **39**, 519–528.

RUTHER, U., and MULLER-HILL, B. (1983). Easy identification of cDNA clones. *EMBO J.* **2**, 1791–1794.

SELSTAM, E., and JACKSON, A. O. (1983). Lipid composition of sonchus yellow net virus. *J. Gen. Virol.* **64**, 1607–1613.

TENINGES, D., and BRAS-HERRENG, F. (1987). Rhabdovirus sigma, the hereditary CO_2 sensitivity agent of *Drosophila:* Nucleotide sequence of a cDNA clone encoding the glycoprotein. *J. Gen. Virol.* **68**, 2625–2638.

TORDO, N., POCH, O., ERMINE, A., KEITH, G., and ROUGEON, F. (1986). Walking along the rabies genome: Is the large G–L intergenic region a remnant gene? *Proc. Natl. Acad. Sci. USA* **83**, 3914–3918.

VAN BEEK, N. A. M., LOHUIS, D., DIJKSTRA, J., and PETERS, D. (1985). Morphogenesis of sonchus yellow net virus in cowpea protoplasts. *J. Ultrastruct. Res.* **90**, 294–303.

WAGNER, R. R. (1987). Rhabdovirus biology and infection. In "The Rhabdoviruses" (R. R. Wagner, Ed.), pp. 9–73. Plenum, New York.

WILEY, D. C., and SKEHEL, J. J. (1990). Viral membranes. In "Virology" (D. N. Fields *et al.*, Eds.), pp. 68–85. Raven Press, New York.

WUNNER, W. H. (1987). Rabies Viruses—Pathogenesis and immunity. In "The Rhabdoviruses" (R. R. Wagner, Ed.), pp. 361–426. Plenum, New York.

ZUIDEMA, D., HEATON, L. A., and JACKSON, A. O. (1987). Structure of the nucleocapsid protein gene of sonchus yellow net virus. *Virology* **159**, 373–380.