Nucleotide Sequences of the mRNA's Encoding the Vesicular Stomatitis Virus N and NS Proteins

CAROL J. GALLIONE,¹ JONATHAN R. GREENE,² LINDA E. IVERSON,^{1,2} AND JOHN K. ROSE^{1*}

Tumor Virology Laboratory, The Salk Institute, San Diego, California 92138,' and Department of Biology, University of California, San Diego, La Jolla, California ⁹²⁰⁹³²

Received 24 March 1981/Accepted 20 April 1981

The complete nucleotide sequences of the vesicular stomatitis virus (VSV) mRNA's encoding the N and NS proteins have been determined from the sequences of cDNA clones. The mRNA encoding the N protein is 1,326 nucleotides long, excluding polyadenylic acid. It contains an open reading frame for translation which extends from the ⁵'-proximal AUG codon to encode ^a protein of ⁴²² amino acids. The N mRNA is known to contain ^a major ribosome binding site at the ⁵' proximal AUG codon and two other minor ribosome binding sites. These secondary sites have been located unambiguously at the second and third AUG codons in the N mRNA sequence. Translational initiation at these sites, if it in fact occurs, would result in synthesis of two small proteins in a second reading frame. The VSV mRNA encoding the NS protein is ⁸¹⁵ nucleotides long, excluding polyadenylic acid, and encodes a protein of 222 amino acids. The predicted molecular weight of the NS protein (25,110) is approximately one-half of that predicted from the mobility of NS protein on sodium dodecyl sulfate-polyacrylamide gels. Deficiency of sodium dodecyl sulfate binding to a large negatively charged domain in the NS protein could explain this anomalous electrophoretic mobility.

The genome of vesicular stomatitis virus (VSV) is ^a single negative strand of RNA about 11,000 nucleotides long (5, 15). The viral N, NS, and L proteins associate with the genomic RNA to form the helical nucleocapsid structure within the enveloped virion (reviewed by Wagner [22]). The transmembrane glycoprotein (G) and the peripheral membrane protein (M) are separable from the nucleocapsid core after disruption of the viral membrane with detergent (22). The L and NS proteins can be dissociated from the nucleocapsid and have been shown to be required for mRNA synthesis from ^a template consisting of N protein and genomic RNA (3). The synthesis, capping, methylation, and polyadenylation of the five VSV mRNA's are probably carried out by the L and NS proteins during transcription from the N protein-genome RNA complex (reviewed by Banerjee et al. [1]). A knowledge of the amino acid sequences of the N and NS proteins should facilitate further structural and functional analyses of these proteins and will be critical to a molecular understanding of their interactions with genomic RNA and other virion components. As part of an analysis of VSV mRNA and protein structure we have determined the nucleotide sequences of the VSV mRNA's encoding the N and NS proteins from

cDNA clones. The amino acid sequences of the proteins are predicted from these sequences. Features of the mRNA and protein sequences are discussed.

MATERIALS AND METHODS

Details of mRNA synthesis, cDNA cloning, and DNA sequencing are provided in the accompanying paper (18).

Primer isolation. Isolation of the primer illustrated in Fig. ³ from pNS173 DNA was as follows. pNS173 DNA (50 μ g) was digested with AluI and treated with alkaline phosphatase, and the 198-basepair blunt-ended AluI fragment (56 base pairs of pBR322 DNA linked to ¹⁴² base pairs of insert DNA) was purified by gel electrophoresis followed by electroelution. The 5'-ends of this fragment were labeled with ^{32}P by using polynucleotide kinase and [γ -³²P]ATP. The fragment was digested with Hinf1, which cuts once and generates staggered ends. The products were then separated by electrophoresis on a sequencing gel (40 cm by 16 cm by 0.35 mm). The 53 nucleotide, single-stranded primer fragment was located by autoradiography and eluted by soaking the gel slice in 0.5 ml of 0.4 M sodium acetate for ⁵ h. The labeled primer fragment separates from the unlabeled complementary strand because it is three nucleotides shorter.

Primer extension and sequencing. Primer derived from approximately 25 μ g of pNS173 DNA (10⁶) dpm) was combined with 100μ g of total VSV mRNA prepared as described previously (18). The mixture of primer DNA and mRNA was then precipitated with ethanol. The pellet was washed with 80% ethanol, dried, and suspended in 10 μ l of 4× reverse transcription buffer containing ²⁰⁰ mM Tris-hydrochloride (pH 8.4), 40 mM $MgCl₂$, 120 mM β -mercaptoethanol, 440 mM KCl, and each of the four dNTP's at ² mM. This mixture was incubated at 40°C for 10 min followed by the addition of 20 μ l of water and 200 U of reverse transcriptase in 10 μ l. Incubation was for 1 h at 42°C followed by phenol extraction and ethanol precipitation. The pellet was resuspended in 100 μ l of 10 mM Tris (pH 7.4) containing ¹⁰ U of RNase A (Worthington Diagnostics) and heated to 90°C for ¹ min. The extended primer was then purified by electrophoresis on a 20% polyacrylamide gel. The gel was subjected to autoradiography, and the largest band (250 nucleotides) was excised, electroeluted, and sequenced.

RESULTS

Determination of the N mRNA sequence. The procedures used for generation of nearly full-length cDNA clones from the VSV N and NS mRNA's are described in the accompanying paper (18). Two cDNA clones designated pN3 and pN4 were used for the N mRNA sequence analysis reported here. The inserts in both of these clones were about 1,350 nucleotides long, including 30 to 40 dG:dC tails at each end, and were flanked by PstI sites. Determination of the nucleotide sequences at the ends of the pN4 insert DNA showed that it contained all but ⁶¹ nucleotides from the ⁵' end of the mRNA and the complete ³' mRNA sequence. This was clear from comparison of the insert DNA sequence with published sequences from both the ⁵' and ³' ends of the N mRNA (11, 13, 19).

The basic strategy for sequencing the N gene was first to isolate the two large HaeIll fragments (Fig. 1) from the insert DNA and then to determine the sequences of the separated strands of each fragment by the Maxam-Gilbert procedure (9). Combined with data published previously (11, 13, 16, 19), these sequences enabled us to establish a tentative sequence for the entire gene. A restriction map was generated from this sequence, and the fragments needed to verify the sequence on both DNA strands were isolated and sequenced (Fig. 1). We also sequenced through all sites used as sequencing endpoints and thus discovered the short sequence between the two central HaeIII sites.

The predicted sequence of the N mRNA is shown in Fig. 2. This 1,326-nucleotide sequence contains a single open reading frame for translation beginning at the AUG in positions ¹⁴ through 16 (within the previously identified ribosome binding site [16]) and ending at positions 1,280 through 1,282. The two other translational reading frames are blocked extensively throughout the sequence. Comparison of the 3'-terminal N mRNA sequences from the VSV Indiana and New Jersey serotypes suggested that the reading frame shown here was correct (12). The predicted N protein sequence is ⁴²² amino acids long with a calculated molecular weight of 47,355.

Determination of the NS mRNA sequence. A cDNA clone of the VSV NS mRNA (pNS173, reference 9) was used for the majority of the sequence analyses reported here. A restriction map of the insert DNA illustrating the sites for restriction enzymes used in the sequence analysis is shown in Fig. 3. To obtain the ⁵' terminal NS mRNA sequence that was not present in the cDNA clone, we isolated a 5^{7} ³²Plabeled DNA primer from the position indicated in Fig. 3 (see above). This primer was then extended to the ⁵' end of the NS mRNA by using reverse transcriptase. The sequence obtained from this elongated primer extended into the ⁵'-terminal sequence of the NS mRNA that was known previously from the sequence of the ribosome binding site (16). This analysis showed that pNS173 lacks ¹²⁶ nucleotides of the ⁵' NS mRNA sequence. An additional clone, pNS319,

FIG. 1. Restriction map of the pNF4 insert DNA. Sites for restriction enzymes that were used in sequencing are shown. The left end corresponds to the 5' end of the N mRNA. Arrows indicate the regions sequenced and the direction of sequencing. Terminal PstI sites are at the junctions of pBR322 sequences with the $dG.dC$ tails. The lengths of the dG:dC tails are not indicated in this figure or in Fig. 3.

r SE; THR1 SE; A A A A^{MET}G SER CAL+ ATE A GAL-CA ASO AND AND AND INS MRINA SEQUENCES 5

CAGTAAT₁₀ AAA A^{MET}G TCT₂ ⁶F1 ATE A GAL-CA^T40 A ^ACA A^TC ATH40G A ^{ARC} A^{SP}C ATHA GA^CF₄ C₆ ²

ARC LYS SER P) LYS LEU PRO ALA ASN GLU ASP PRO VAL GLU TYR PRO ALA ASP TYR PHE ARG LYS SER LYS
AAAAC TICCQGAA CORGAT CCAGT GGAA TACCCGGCGAGT TACTT CAAA AT CAAA GAGATT CCTCTTTACATCAN THA LUS ARAGTTGTGRA.
150 140 150 140 150 140 150 140 150 140 150 140 150 160 \cdot $\begin{bmatrix} 1.6 & 6 & \text{A} & \text{T} & \text{T} & \text{A} & \text{C} & \text{T} & \text{T} & \text{T} & \text{T} & \text{A} & \text$ E)i GLY LEU LYS SER GLY ASN VAL SER ILE ILE HIS VAL ASN SER TYR LEu TYR GLY ALA ^C CAL :C .C: L-L:A ^s - ^C^G ^G ^A AAT ^G TAT CA AT ^C ^A TA CAT ^G ^T ^C ^A ^A ^C AG ^C TACT ^T ^G TAT ^G GAG ^C 190 ~~~~200 ²¹⁰ ²²⁰ ²³⁰ ²⁴⁰ LEU LYS ASP ILE ARG GLY LYS LEU ASP LYS ASP TRP SER SER PHE GLY ILE ASN ILE GLY
ATTAAAGGACATCCGGGGTAAGTTGGATAAGATTGGTCAAGTTTCGGAATAAACATCGG LYS ALA GLY ASP THR ILE GLY ILE PHE ASP LEU VAL SER LE LYS ALA LEU ASP GLY VAL GAAAGCAGGGGATACAATCGGAATATTTGACCTTGTATCCTTGAAAGCCCTGGACGGCGT ³¹⁰ ³²⁰ ³³⁰ ³⁴⁰ ³⁵⁰ ³⁶⁰ LEUT PRO ASP GLY VAL SER ALA SER ARG 11HR SER ALA AGP AS ALA ASP LYS TRP LEU PRO AGINAL
CTTCCAGATGGAGTATCGGATGCTTCCAGAGCGGAGGAGAACCAGCGGATGACAATGGTTGCCTTC
420 400 410 420 TYR LEU LEU GLY LEU TYR ARG VAL GLY ARG THR GLAN MET PRO GLU TYR ARG LYS LEU GTATCTACTTATACAGAGAGTGGGCAGAGACA
GTATCTACTTGGCTTATACAGAGTGGGCAGACACAAATIGCCTGAATACAGAAAAAGCT MET ASP GLY LEU THR ASN GLIN CYS LYS MET ILE ASN GLU GLIN PHE GLU PRO LEU VAL PRO
CATGGATGGGCTGACAAATGCAAAATGAAATGAATGAACCOTTTGAACCOTCTTGTGC GLU GLY ARG ASP ILE PHE ASP VAL TRP GLY ASN ASP SER ASN TYR THR LYS ILE VAL ALA
AGAAGGTCGTGGGACATTTTTTTGATGTGGGGGAAATGACAGTAATTACACAGAAAATTGTCGC ALA VAL ASP MET PHE PHE HIS MET PHE LYS LYS HIS GLU CYS ALA SER PHE ARG TYR GLY
TGCAGTGGACATGTTCTTCCACATGTTCAAAAACATGAATGTGCCTCGTTCAGATACGG THR ILE VAL SER ARG PHE LYS ASP CYS ALA ALA LW ALA THR PHE GLY HIS LW CYS LYS AACTATTGTTTCCAGATTCAAAGATTGTGCTGCATTGGCAACATTTGGACACCTCTGCAA ⁶⁷⁰ ⁶⁸⁰ ⁶⁹⁰ ⁷⁰⁰ ⁷¹⁰ ⁷²⁰ ILE THR GLY MET SER THR GLU ASP VAL THR THR TRP ILE LEU ASN ARG GLU VAL ALA ASP AGATAACCGGAAGAAGAAGAAGAAGAAGA
AATAACCGGAATGTCTACAGAAGATGTAACCGAGGATCGTGATIGAACCGAGAAGTTGCAGA GLU MET VAL GLN MET MET LEU PR GLY GLN GLU ILE ASP LYS ALA ASP SER TYR MET PR) TGAAATGGTCCAAATGATGCTTCCAGGCCAAGAAATTGACAAGGCCGATTCATACATGCC ⁷⁹⁰ ⁸⁰⁰ ⁸¹⁰ ⁸²⁰ ⁸³⁰ ⁸⁴⁰ TYR LWU ILE ASP PHE GEY LWU SER ESR LYS SER PR) TYR SER SER VAL LYS ASN PR) ALA TTATTTGATCGACTTTGGATTGTCTTCTAAGTCTCCATATTCTTCCGTCAAAAACCCTGC 850 860 870 880 890 900 PHE HIS PHE TRP GLY GLN LEU LATA LEU LEU LEU LOI PARG ARG ALA ARG ASN ALA
CTTCCACCTTCTGGGGGCAATTGACCAGCTCTCTCAGGAGCTGCAGAATCCAGGGAATCCAGGGGGGGAATUW
60 CAAGGABAATUW90 CAAGCAGCAGA9 OLIHAA ARG GLN PR) ASP ASP ILE EL TYR THR SER LW TH THR ALA GLY LS LW TYR ALA TYR CCGACAGCCTGATGACATTGAGTATACATCTCTTACTACAGCAGGTTTGTTGTACGCTTA ⁹⁷⁰ ⁹⁸⁰ ⁹⁹⁰ ¹⁰⁰⁰ ¹⁰¹⁰ ¹⁰²⁰ ALA VAL GLY SER SER ALA ASP LEU ALA GLN GLN PHE CYS VAL GLY ASP ASN LYS TYR THR
TGCAGTAGGACTTGCCGCGCGCGACTAGAACAGTTTTGGTGGGGGAGATACAAATAC
1030 1030 1040 1040 1050 1050 1050 1060 1070 1070 1080 1090 1090 1080 1070 1080 1070 PRO ASP ASP SER THE GLY GLY LEU THE GAC AG AG COCOCACTAG SP SER
TCCAGATGATAGTACCGGAGGATTGACGATTGATGCATGGCAGGAGGAGAGGAGAGGATGTGGT
1120 1130 1130 1140 GEU TRP LW GLY TRP PHE GLU ASP GLN ASN ARG LYS PR THR PR) ASP MET MET GLN TYR CGAATGGCTCGGATGGTTTGAAGATCAAAACAGAAAACCG8ACTCC TGATATGATGCAGT ^A 1150 1160 1170 1180 1190 1200 ALA LYS ARG ALA VAL MET SER LE GEN GLY LW ARG GEL LYS THR ILE GLY LYS TYR ALA TGCGAAAAGAGCAGTCATGTCACTGCAAGGCCTAAGAGAGAAGACAATTGGCAAGTATGC ¹²¹⁰ ¹²²⁰ ¹²³⁰ ¹²⁴⁰ ¹²⁵⁰ ¹²⁶⁰ LYS SER GLU PHE ASP LYS *** TAAGTCAGAATTTGACAAATGACCCTATAATTCTCAGATCACCTATTATATATTATGCTA ¹²⁷⁰ ¹²⁸⁰ ¹²⁹⁰ ¹³⁰⁰ ¹³¹⁰ ¹³²⁰

CATATG

sequence except for 32 nucleotides from the 5'

was isolated which has the complete NS mRNA end. The 5' sequence determined from this clone sequence except for 32 nucleotides from the 5' is indicated in Fig. 3.

FIG. 2. Nucleotide sequence of the VSV mRNA encoding the N protein and the predicted N protein sequence. The shaded nucleotide sequences represent the three ribosome binding sites in the N mRNA.

FIG. 3. Restriction map of the pNS173 insert DNA. Sites for restriction enzymes that were used in sequencing are shown. The heavy line indicates the DNA segment used as ^a primer on the NS mRNA to obtain the ⁵'-terminal mRNA sequence.

DISCUSSION

Translation of the N mRNA. The nucleotide sequence presented here for the VSV mRNA encoding the N protein predicts ^a protein of ⁴²² amino acids initiated at the first AUG codon in the mRNA (Fig. 2). Our previous analysis of ribosome-protected initiation sites in N mRNA showed that three separate sites were protected within the sequence (16). The major site occurred at the ⁵'-proximal AUG codon and contained about 75% of the protected material. The minor sites were each found to contain 10 to 15% of the protected material. Partial sequence analysis showed that both minor sites contained single AUG codons. The complete sequence of the N mRNA contains ⁴⁵ AUG codons in all three frames. We compared the partial sequences of the T1 oligonucleotides
from the minor ribosome binding sites with the from the minor ribosome binding sites with the sequences around each AUG codon. These partial sequences can be aligned perfectly at the second and third AUG codons in the mRNA (Fig. 4), but not at any of the other AUG codons. An assignment of one of these sites to the second AUG codon was made previously on the basis of partial sequence information for the N mRNA (17).

The generally accepted view of eucaryotic protein synthesis initiation (reviewed in reference 8) is that eucaryotic ribosomes initiate only at single sites on mRNA's. These sites are usually, but not always, the first AUG in the mRNA (8). The fact that ribosomes bind only at the three ⁵'-proximal AUG codons in the N mRNA rather than at any of the ⁴² other AUG codons suggests that this binding results from "entry" of ribosomes at the ⁵' end of the mRNA. Presumably, features of the mRNA structure (such as distance from the ⁵' end) determine the extent of binding at the three sites. These multiple binding sites could be artifacts of the in vitro binding

B

$$
\begin{array}{cc} 170 & 180 & 190 \\ \text{ATCTAAGAGGATATCTCTACCAAGGCCCTCAATCCGG} \\ ----YAG & -YAUG \\ \text{AG (G)} & ----YAG (G) \end{array}
$$

FIG. 4. Alignment of Tl oligonucleotide sequences with N mRNA sequences. Partial sequences of two secondary ribosome binding sites from the N mRNA (16) are shown aligned with the appropriate portions of the N mRNA sequence. Y indicates ^a pyrimidine, and parentheses indicate nearest neighbors. Sizes of the Tl oligonucleotides predicted are consistent with the migration of each oligonucleotide on the twodimensional homochromatography system (16). Numbers are relative to the ⁵' end of the N mRNA.

conditions, although we did not observe secondary binding sites in other VSV mRNA's under identical conditions. If ribosomes initiated translation at these sites, only small proteins of 30 and 23 amino acids would be produced. Such small proteins would have gone undetected previously. Because the amino acid sequences of these putative proteins can be predicted from the N mRNA sequence, it may be possible to detect them by labeling VSV-infected cells with the appropriate amino acids.

NS mRNA and protein structure. The nucleotide sequence predicted here for the VSV mRNA encoding the NS protein is ⁸¹⁵ nucleotides long, excluding polyadenylic acid. It encodes a protein of 222 amino acids (25,110 daltons) initiated from the ⁵'-proximal AUG codon in the mRNA (Fig. 4). Previous estimates of the molecular weight of NS from its migration on sodium dodecyl sulfate-polyacrylamide gels range from 39,000 to 50,000 (see reference 7). This discrepancy suggests that NS has an unusual monomeric structure or that it perhaps migrates as ^a dimer. Phosphorylation of NS cannot explain its anomalous migration because removal of phosphate residues from NS decreases rather than increases its mobility on sodium dodecyl sulfate-polyacrylamide gels (4). Inspection of the distribution of charged residues within the predicted NS protein sequence does reveal the unusual feature of a large domain containing 18 negatively charged residues and no positively charged residues (Fig. 5). A deficiency of sodium dodecyl sulfate binding to this negatively charged domain might explain the

anomalous electrophoretic mobility. Codon usage. As found for the G and M mRNA's (18), there is also a deficiency in the CG dinucleotide in the N and NS mRNA's. One would predict that 58 CGs would occur at random in the N mRNA coding region and ³³ would occur in the NS coding region, whereas only ³⁰

and 16, respectively, are found. The codon usage for both mRNA's is shown in Fig. 6. In the N mRNA there is ^a clear bias against arginine codons of the form CGN and against NCG codons $(N = A, G, C, or T)$. A similar bias is also seen between adjacent codons where 31 junctions of the type NNC-GNN would be expected at random, whereas only ¹⁸ are found. In the NS mRNA there is no bias against the CGN codons, but the NCG codons are deficient. Codon junctions NNC-GNN are strongly deficient in NS, as only 3 are found where 14 would be expected at random. Deficiency in the CG dinucleotide may prove to be a general feature in the sequences of RNA genomes of eucaryotic viruses as it has been observed in the mRNA encoding an influenza virus hemagglutinin (14).

Remainder of the genome. The sequences presented here, combined with the previously determined leader RNA (2), ³'-terminal genomic sequences (6, 11, 20), and intergenic sequences (10, 17), provide a continuous sequence of 4,723

10, 17), provide a continuous sequence of 4,723
 C^{RG} GW_G TR_{AG} C_{RC} L_R_G C_{RC} C_G_C C_G_C C_G_C C_G_C C_G_C ASP GA) ALA VAL GLY GLU ILE ASP GW ILE GW ALA GM ARG ALA GA LYS ^S ASM TYR GGATCAGGCGGTAGGAGAGATAGATGAGATCGAAGCACAACGAGCTGAAAAGTCCAATTA ⁷⁰ ⁸⁰ ⁹⁰ ¹⁰⁰ ¹¹⁰ ¹²⁰ GW LWU PHE GM GA) ASP GLY VAL GAU GAW HIS THR PALYS ⁰ SER TYR PHE GM ALA ALA TGAGTTGTTCCAAGAGGATGGAGTGGAAGAGCATACTAAGCCCTCTTATTTTCAGGCAGC ¹³⁰ ¹⁴⁰ ¹⁵⁰ ¹⁶⁰ ¹⁷⁰ ¹⁸⁰ NW SR¶¶ aR St AW: PRO CHR ILE CM AN AG GGY LEU TYR ALA GLN AGATGATTCGWA CACAGAATCTGAVACCA GAkAATTG ^A AGCAATCAAGGTTTGTATGCA CA ¹⁹⁰ ²⁰⁰ ²¹⁰ ²²⁰ ²³⁰ ²⁴⁰ ASOP PRO GUAT ALA GUAT GUN VAL GUAT GLY PHE ILE GUN GLY PRO LEU ASOP ARE TRA ASTROLA AG
GGATCCAGAAGCTGAGCAAGTTGAAGGCTTTATACAGGGCCTTTAGATGCAGAGO
200 290 200 270 200 270 200 200 270 270 280 290 290 290 290 290 290 290 290 29 GLO GAL MAN VAL AND THE THR SER MAN TRE LYS GIN PRO GLU LEU GLU SER ASP GLU
TGAGGAAGTGGAATCTTGTATTACTTGGGAAACAGCCTGAGCTTGAGTGAATCTGAG
300 330 330 330 330 340 340 340 350 350 350 350 350 350 350 350 350 360 360 360 360 360 HIS GLY LYS T*R LEU ARG LWU THR SER PRO GLU GLY LEU SER GLY GA) GLN LYS SER GLN GCATGGAAAGACCTTACGGTTGACATCGCCAGAGGGTTTAAGTGGAGAGCAGAAATCCCA ³⁷⁰ ³⁸⁰ ³⁹⁰ ⁴⁰⁰ ⁴¹⁰ ⁴²⁰ TRP LEU SER THR ILE LYS. ALA VAL VAL GMN SER ALA LYS TYR TRP ASN LWU ALA GLU CYS GTGGCTTTCGACGATTAAAGCAGTCGTGCAAAGTGCCAAATACTGGAATCTGGCAGAGTG 430 440 450 460 470 480 The $\tilde{\bf r}$ bitch $\tilde{\bf r}$ gets $\tilde{\bf r}$ gan $\tilde{\bf r}$ of $\tilde{\bf r}$, which it is a limit in the value of $\tilde{\bf r}$ and $\tilde{\bf r}$ in $\tilde{\bf r}$ and $\tilde{\bf r}$ in $\tilde{\bf r}$ and $\tilde{\bf r}$ is a limit $\tilde{\bf r}$ and $\tilde{\bf r}$ GTGGCTTTGACGATTAAAGCAGTCGTGGAAAGTCCCAAATACTGGAATATGCAGAGT
THE PHE GUI ALA SER GUY GUI GUY VAL ILE MET LYS GU ARG GUN ILE THE PRO ASP VAL
CACATTTC490 6CATCGGG900 AAGGGGTC 510
SACATTTC490 540 TYR LYS VAL THR PRO VAL MET ASN THR HIS PRO PHE PRO ILE ARG SER SER ILE ARG CYS
'ATAAGGTCACTCCAGTGAATGAACACACAGTCCGTTCCCAATCAGAAGCAGTATCAGAGO
000 550 550
THE PRO VAL MET ASN THE HIS PRO PHE PRO ILE ARG SER SER ILE ARG CY
TCACTCCAGTGATGAACACAGATCCGTTCCCAATCAGAAGGAGTATCAGATG
550 590 590 560
550 590 590 590 500 LEU VAL SER LEU LYS ASP ILE HIS ASP PHE PRO THR GLN GLU SER LYS SER SER ALA SER
TTTGGTCTTCTAGAGACAAGTCTTCATGACTTTCAACCCAAGGOAGAGAAGCAAGTCTTCAGCC
660 630 640 630 640 650 650 HIS HIS ILE L (LY *** ^T ^C AC CATAT CCTT GGA TGA AT ^T GT TCTCAT CTAG AG GAGAG ^T TCAT CTCT ^G ^T CGGAG GTO 670 680 690 700 710 72& ACGGACGA₁₃TGTCTCATA₂AGAGGCCAT 5C₆TGCTCGGC 5TGGAGATACA₇₇₀AAAGTTGTA 5A ATCAGGCGAGAGTCAAATATTCTCTGTAGACTATG

FIG. 5. Nucleotide sequence of the VSV mRNA encoding the NS protein, and the predicted NS protein sequence. The shaded nucleotide sequence indicates the ribosome binding site. Negatively charged amino acids in the negatively charged domain are shaded.

nucleotides from the ³' half of the VSV genome. Using the value of 11,278 nucleotides calculated by Repik and Bishop (15) for the size of the genome, this leaves 6,496 nucleotides to encode the L mRNA which terminates ⁵⁹ nucleotides from the ⁵' end of the genome (21). This sequence is more than sufficient to encode the L protein, which is estimated to have a molecular weight of about 190,000 (22).

ACKNOWLEDGMENTS
We thank Jim Ostlund, Joe Zwass, and Jeff Velten for help We thank Jim Ostiund, Joe Zwass, and Jeff Velten for help
with computer analysis of sequence data and the members of the Tumor Virology Laboratory for their helpful suggestions. We are grateful to Carolyn Goller for typing this and the accompanying manuscript.
This work was supported by Public Health Service grants

AI-15481 from the National Institute of Allergy and Infectious AI-15481 from the National Institute of Allergy and Infectious Diseases and CA-14195 from the National Cancer Institute.

LITERATURE CITED

- 1. Banerjee, A. K., G. Abraham, and R. J. Colonno. 1977. Vesicular stomatitis virus: mode of transcription.
J. Gen. Virol. 34:1-8.
- 2. Colonno, R. J., and A. K. Banerjee. 1978. Complete 2. Colonno, R. J., and A. K. Banerjee. 1978. Complete
nucleotide sequence of the leader RNA synthesized in vitro by vesicular stomatitis virus. Cell 15:93-101.
3. **Emerson, S. U., and Y. H. Yu.** 1975. Both NS and L.
- 3. **Emerson, S. U., and Y. H. Yu.** 1975. Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. J. Virol. 15:1348-1356.
4. Hsu, C. H., and D. Kingsbury. 1980. Vesicular stomatitis
- virus morphogenesis is accompanied by covalent protein modifications, p. 613-622. In B. N. Fields, R. Jaentein modifications, p. 613-622. In B. N. Fields, R. Jaen-
isch, and C. F. Fox (ed.), Animal virus genetics. Academic Press, Inc., New York.
5. Huang, A. S., and R. R. Wagner. 1966. Comparative
- 5. Huang, A. S., and R. R. Wagner. 1966. Comparative sedimentation coefficients of RNA extracted from plaque-forming and defective particles of vesicular stomatitis virus. J. Mol. Biol. 22:381-384.
- 6. Keene, J. D., M. Schubert, and R. A. Lazzarini. 1980. Intervening sequence between the leader region and the nucleocapsid gene of vesicular stomatitis virus RNA. J.
Virol. 33:789-794.
- Virol. 33:789-794. 7. Knipe, D., J. K. Rose, and H. F. Lodish. 1975. Translation of individual species of vesicular stomatitis viral mRNA. J. Virol. 15:1004-1011.
- mRNA. J. VIrol. **15:1004-1011**.
8. **Kozak, M.** 1978. How do ribosomes select initiation regions in mRNA? Cell 15:1109-1123.
- 9. Maxam, A. M., and W. Gilbert. 1980. Chemical sequencing of DNA. Methods Enzymol. 65:501-561.
- 10. McGeoch, D. J. 1979. Structure of the gene N:gene NS intercistronic junction in the genome of vesicular sto-
matitis virus. Cell 17:673-681.
- 11. McGeoch, D. J., and A. Dolan. 1979. Sequence of 200 nucleotides at the ³'-terminus of the genome RNA of vesicular stomatitis virus. Nucleic Acids Res. 6:3199-3211.
- 12. McGeoch, D. J., A. Dolan, and C. R. Pringle. 1980. Comparisons of nucleotide sequences in the genomes of the New Jersey and Indiana serotypes of vesicular
- 13. McGeoch, D. J., and N. Turnbull. 1978. Analysis of the ³'-terminal sequence of vesicular stomatitis virus N protein mRNA. Nucleic Acids Res. 5:4007-4024.
14. Porter, A. G., C. Barber, N. H. Carey, R. A. Hallewell,
- 14. Porter, A. G., C. Barber, N. H. Carey, R. A. Hallewell, G. Threlfall, and J. S. Emtage. 1979. Complete nucleotide sequences of an influenza virus haemagglutinin gene from cloned DNA. Nature (London) 282:471-477.
- 15. Repik, P., and D. H. L. Bishop. 1973. Determination of the molecular weight of animal RNA viral genomes by nuclease digestions. J. Virol. 12:969-983.
- 16. Rose, J. K. 1978. Complete sequences of ribosome recognition sites in VSV mRNAs. Cell 14:345-353.
- 17. Rose, J. K. 1980. Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus.
Cell 19:415-421.
- 18. Rose, J. K., and C. J. Gallione. 1981. Nucleotide sequences of the mRNA's encoding the vesicular stomatitis virus G and M proteins determined from cDNA clones containing the complete coding regions. J. Virol. 39:519-528.
- 19. Rose, J. K., and L. E. Iverson. 1979. Sequences from the ³' ends of vesicular stomatitis virus mRNA's as determined from cloned DNA. J. Virol. 32:404-411.
- 20. Rowlands, D. J. 1977. Sequences of vesicular stomatitis virus RNA, N protein mRNA, and their junction. Proc. Natl. Acad. Sci. U.S.A. 76:4793-4797.
- 21. Schubert, M., J. D. Keene, R. C. Herman, and R. A.

Lazzarini. 1980. The site on the VSV genome specifying polyadenylation and the end of the L gene mRNA. J. Virol. 34:550-559.

22. Wagner, R. R. 1975. Reproduction of rhabdoviruses, p. 1-93. In H. Fraenkel-Conrat and R. R. Wagner (ed.), Comprehensive virology. Plenum Publishing Corp., New York.