JOURNAL OF VIROLOGY, Aug. 1981, p. 519–528 0022-538X/81/080519-10\$02.00/0

Nucleotide Sequences of the mRNA's Encoding the Vesicular Stomatitis Virus G and M Proteins Determined from cDNA Clones Containing the Complete Coding Regions

JOHN K. ROSE* AND CAROL J. GALLIONE

Tumor Virology Laboratory, The Salk Institute, San Diego, California 92138

Received 24 March 1981/Accepted 20 April 1981

The complete nucleotide sequences of the vesicular stomatitis virus mRNA's encoding the glycoprotein (G) and the matrix protein (M) have been determined from cDNA clones that contain the complete coding sequences from each mRNA. The G protein mRNA is 1,665 nucleotides long, excluding polyadenylic acid, and encodes a protein of 511 amino acids including a signal peptide of 16 amino acids. G protein contains two large hydrophobic domains, one in the signal peptide and the other in the transmembrane segment near the COOH terminus. Two sites of glycosylation are predicted at amino acid residues 178 and 335. The close correspondence of the positions of these sites with the reported timing of the addition of the two oligosaccharides during synthesis of G suggests that glycosylation occurs as soon as the appropriate asparagine residues traverse the membrane of the rough endoplasmic reticulum. The mRNA encoding the vesicular stomatitis virus M protein is 831 nucleotides long, excluding polyadenylic acid, and encodes a protein of 229 amino acids. The predicted M protein sequence does not contain any long hydrophobic or nonpolar domains that might promote membrane association. The protein is rich in basic amino acids and contains a highly basic amino terminal domain. Details of construction of the nearly full-length cDNA clones are presented.

Vesicular stomatitis virus (VSV) buds from the surface of the host cell and thereby acquires a membrane that contains spikes composed of the single viral glycoprotein (G). Our previous studies have shown that G protein has a COOHterminal basic domain of 29 amino acids which is internal to the lipid bilayer of the virion and an adjacent hydrophobic segment of 20 amino acids which spans the bilayer (27). The NH₂terminal 95% of G is external to the lipid bilayer and contains two asparagine-linked complex oligosaccharides (10, 21). G plays an essential role in binding of the virus to the host cell and presumably plays a role in directing budding of virus from the plasma membrane of the host cell (2, 5).

The G protein is synthesized on membranebound polyribosomes (18) and inserted into the rough endoplasmic reticulum as a nascent protein chain (29, 34). A short hydrophobic signal sequence of 16 amino acids is cleaved from the NH₂ terminus after insertion into the rough endoplasmic reticulum (12, 15, 22). G assumes a transmembrane configuration in the rough endoplasmic reticulum (6, 13) with only a small COOH-terminal segment exposed on the cytoplasmic face of the rough endoplasmic reticu-

DOCKE

lum. Transport of G to the plasma membrane occurs via the Golgi apparatus (1a), with interorganelle transport probably occurring via coated vesicles (28). At a late stage of transport to the plasma membrane, one or two molecules of fatty acid are esterified to G (31).

The VSV matrix protein (M) is thought to be a peripheral membrane protein that lines the inner surface of the virion envelope, perhaps interacting with the lipid bilayer, the internal portion of G, and the nucleocapsid core (reviewed by Wagner [35]). In addition to a structural function, M plays a role in directing budding of virus from infected cells (14) and may be involved in regulating transcription of mRNA from the single negative strand of genomic RNA (4, 7, 9, 16).

In contrast to G, M is synthesized on free polyribosomes (18) and associates rapidly with the plasma membrane fraction after synthesis (14). The nature of the association of M with the plasma membrane fraction is not clear. It could be by association with the lipid bilayer in regions containing G protein or by association with other proteins such as those in nucleocapsids which are already associated with the plasma membrane (14).

close corre-

Vol. 39, No. 2

520 ROSE AND GALLIONE

Knowledge of the complete primary amino acid sequences of G and M proteins is clearly critical to a molecular understanding of their interactions with membranes and other cellular and viral components. Our previous studies have employed VSV cDNA clones that contain only fractions of the coding sequences from each mRNA (24, 27), and only terminal mRNA and protein sequences were reported. We report here the isolation and the complete nucleotide sequences of cDNA clones that contain the entire coding sequences of G and M protein mRNA's. From these sequences we predict the amino acid sequences of the G and M proteins and discuss features of these predicted sequences.

MATERIALS AND METHODS

Materials. Reverse transcriptase was supplied by J. Beard, St. Petersburg, Fla. The Klenow fragment of DNA polymerase I and most restriction endonucleases were purchased from New England Biolabs, Beverly, Mass. Nuclease S1, *Pst*I, and T4 polynucleotide kinase were from Boehringer Mannheim, Indianapolis, Indiana, and terminal deoxynucleotidyl transferase was from Bethesda Research Laboratories, Bethesda, Md. Oligo(dT)₁₂₋₁₈ and oligo(dT) cellulose (T3) were from Collaborative Research, Waltham, Mass. [α^{-32} P]dCTP and dGTP were from Amersham/Searle, Chicago, Ill., and unlabeled dNTP's were from P-L Biochemicals, Milwaukee, Wis.

RNA synthesis. VSV mRNA was synthesized in vitro in a 10-ml reaction exactly as described previously (26), with the following modification. After 4 h the reaction was stopped by addition of sodium dodecyl sulfate and sodium acetate to final concentrations of 1% and 0.5 M, respectively. The entire mixture was then passed through a column containing 0.5 g of oligo(dT)-cellulose, and the column was washed with 10 ml of 0.4 M sodium acetate. The bound mRNA was eluted with distilled water and precipitated with ethanol. All mRNA was from the San Juan strain of the Indiana serotype of VSV.

First-strand DNA synthesis and purification. First-strand DNA copies of the mRNA's encoding the N, NS, M, and G proteins were synthesized in a 4-ml reaction containing 300 µg of total VSV mRNA, 0.5 mM dATP, dGTP, and dTTP, 0.25 mM [α -32P]dCTP (10 Ci/mmol), 30 mM $\beta\text{-mercaptoethanol},$ 120 mM KCl, 10 mM MgCl₂, 4,000 U of reverse transcriptase, 300 μg of $oligo(\bar{d}T)_{12\text{--}18}$ and a cytoplasmic extract (1 mg of total protein) from baby hamster kidney (BHK-21) cells prepared as described below. After incubation for 30 min at 42°C, the reaction mixture was extracted with phenol. Unincorporated dNTP's were separated from cDNA by chromatography of the aqueous phase on Sephadex G-50. The excluded fraction was lyophilized in a silicated glass tube. The yield of total cDNA from 300 µg of VSV mRNA was approximately 40 µg as determined from the incorporation of $[\alpha^{-32}P]dCTP$. The full-length cDNA copies of each mRNA species were purified by electrophoresis on a 1.5% alkaline agarose gel (30 mM NaOH, 5 mM

DOCKE.

EDTA; 1.5 mm by 18 cm by 20 cm). The cDNA's were located by autoradiography (Fig. 1) or by staining with ethidium bromide and UV illumination. Stained bands were excised from the gel, electroeluted, and precipitated with ethanol. The yield of full-length cDNA's was approximately 0.7 μ g of G, 2 μ g of N, and 3 μ g of both NS and M. Reaction conditions given were optimized with respect to the concentrations of reverse transcriptase, KCl, and cell extracts to give maximal sizes of cDNA's.

Cell extract preparation. BHK cells $(8 \times 10^8$ cells) growing in a 1-liter suspension were pelleted by centrifugation and suspended in 1.5 times the packed cell volume of 10 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) (pH 7.5), 10 mM KCl, 1.5 mM magnesium acetate, and 7 mM β -mercaptoethanol. Cells were broken by 50 strokes of a Dounce homogenizer and centrifuged at 15,000 × g for 20 min. The supernatant was dialyzed for 16 h against 1 liter of 10 mM HEPES (pH 7.5), 90 mM KCl, 1.5 mM magnesium acetate, and 7 mM β -mercaptoethanol. The total protein concentration in the extract was 15 to 20 mg/ml. All steps were carried out at 4°C, and samples of the extract were stored at -80° C.

Second-strand cDNA synthesis. Reactions for second-strand DNA synthesis (200 µl) included 100 mM HEPES buffer (pH 6.9), 10 mM MgCl₂, 60 mM KCl, 1 mM dATP, dCTP, dGTP, and dTTP, 0.5 µg of a purified cDNA, and 5 U of the Klenow fragment of DNA polymerase I. Reactions were for 15 h at 15°C and were terminated by extraction with phenol followed by precipitation with ethanol. Samples were then resuspended in 75 μ l of a solution containing 3 mM ZnCl₂, 30 mM sodium acetate (pH. 4.5), 300 mM NaCl, and 75 U of S1 nuclease. Digestion was for 30 min at 37°C. S1-treated DNAs were extracted with phenol and precipitated with ethanol and then purified by electrophoresis on a 6% polyacrylamide gel. The nearly full-length double-stranded DNAs were detected by autoradiography or staining with ethidium bromide. The amounts of double-stranded DNAs recovered were 100 ng (G size class), 180 ng (N size class), and 450 ng (NS and M size class), as calculated from the cpm of ^{32}P in the first-strand DNA. Doublestranded DNAs were electroeluted and precipitated with ethanol.

Homopolymer addition and cloning. Homopolymer addition was by a modification of a published procedure (30). Addition of dCMP residues to doublestranded cDNA was for 10 to 30 min at 37°C in a 25- μl reaction containing 100 to 500 ng of DNA, 0.2 mM dCTP, and an amount of terminal deoxynucleotidyl transferase (Tdt) that was known to add approximately 10 to 20 dC residues to HaeIII fragments (5 pmol of total ends) of pBR322 DNA. The number of dC residues added was calculated from the incorporation of $[\alpha^{-32}P]dCTP$ (100 Ci/mmol) that was included in test reactions. Units reported by the supplier were generally 10- to 50-fold greater than what we observed. Addition of 10 to 20 dG residues to PstIcleaved pBR322 DNA was carried out as above (0.2 mM dGTP) using 1 to 2 μ g of DNA. The linear form of PstI-cleaved pBR322 was always purified by agarose gel electrophoresis and electroelution before tailing.

J. VIROL.

Vol. 39, 1981

Equimolar quantities of dC-tailed insert DNA and dGtailed plasmid DNA were annealed for 15 min at 4°C in 10 µl of 0.5 M NaCl-10 mM Tris (pH 7.4) and used to transform Escherichia coli strain C600 to tetracycline resistance (30). From 10 to 40 colonies were obtained per ng of insert DNA. Approximately 80% of the colonies obtained were ampicillin sensitive and had inserts at the PstI site. Small preparations of plasmid DNA (1 to 2 μ g) were analyzed for the size of the insert DNA by PstI digestion. HaeIII and HinfI digestions gave partial restriction maps for each insert. By comparing these maps with previous data (24, 25, 27) we were able to assign the clones unambiguously to the appropriate mRNA's. Although apparently fulllength double-stranded DNAs were used as starting material, a large fraction (80% for G) of the inserts had deletions at one or both ends.

DNA sequence analysis. All sequence analysis was by the Maxam-Gilbert procedure as described previously (17, 25). DNA fragments (25 to 50 pmol) were prepared by restriction enzyme digestion, alkaline phosphatase treatment, polyacrylamide gel electrophoresis, and electroelution (25). End labeling was with 500 μ Ci of [γ -³²P]ATP and 2 U of polynucleotide kinase (24). Labeled DNA strands were separated by electrophoresis on thin 5% polyacrylamide gels (0.35 mm by 16 cm by 40 cm) rather than the thicker gels suggested by Maxam and Gilbert (17). Thin gels gave better resolution of closely spaced DNA strands. Gels were run at 1,200 V with an electric fan for cooling, and strand separation of fragments up to 800 base pairs long was accomplished by using electrophoresis times of less than 8 h. Occasionally fragments for sequencing were obtained by cleavage with a second restriction enzyme followed by gel electrophoresis. Sequencing gels (0.35 mm by 16 cm by 40, 85, or 152 cm) were electrophoresed at 1,500 to 6,000 V, such that the gel remained warm (>40°C) throughout the run. Reaction times were reduced substantially for samples on which the sequence was to be read for more than 300 nucleotides from the labeled end. Generally 100,000 to 500,000 dpm of ³²P-labeled fragment were loaded per lane on 152-cm gels so that exposure could be carried out at high resolution without fluorescent screens. Long (85 or 152 cm) gels were cut in sections, transferred to old exposed sheets of X-ray film for backing, covered with plastic wrap, and exposed to Kodak XR-5 film for 1 to 4 days.

RESULTS

The basic strategy we employed to obtain nearly full-length cDNA clones was: (i) to optimize reverse transcription of full-length firststrand DNA copies of each mRNA; (ii) to use isolated, full-length first strands as templates for second-strand DNA synthesis; and (iii) to purify the nearly full-length double-stranded DNAs (obtained after S1 nuclease treatment) by gel electrophoresis before cloning. All cloning was of dC-tailed inserts into the dG-tailed *Pst*I site of pBR322 (3).

In previous experiments we and others found that addition of a crude cell extract greatly enhanced the synthesis of full-length VSV mRNA's by the VSV virion transcriptase (1, 26). Because our initial reverse transcription experiments showed that the yields of full-length cDNA's of VSV G and N mRNA's were low relative to the yields of cDNA's of the smaller mRNA's (NS and M), we examined the effect of such an extract on reverse transcription. Figure 1 shows the products of reverse transcription of total VSV mRNA synthesized in the absence and presence of the extract. The yields of fulllength G and N cDNA's were increased 5- and 2-fold, respectively, whereas the M and NS cDNA's were increased only marginally (1.2fold). Because of the dramatic enhancement of the yield of large cDNA, the cell extract was included in all syntheses of single-stranded cDNA's for cloning. The mechanism of action of the cell extract is not known, but it presumably acts by inhibiting RNase or by promoting unfolding of mRNA secondary structure.

G and M cDNA clones contain the complete coding sequences. Two recombinant plasmids, pG1 and pM309, were obtained which had insert sizes of ca. 1,700 and 850 nucleotides, respectively, consistent with their having nearly complete copies of the G and M mRNA sequences. To determine the exact 5' endpoints of each cloned sequence relative to the mRNA, we determined the nucleotide sequence at the end of the insert corresponding to the 5' end of the RNA. The sequencing gel showing this region from the pG1 insert is shown in Fig. 2. The sequence preceding the homopolymer tails is



FIG. 1. Reverse transcripts of VSV mRNA synthesized in the presence (lane 2) or absence (lane 1) of a cytoplasmic cell extract. Reverse transcripts were labeled with $[\alpha^{-32}P]$ dCTP and analyzed by electrophoresis on an alkaline agarose gel. Identities of the cDNA bands in the autoradiogram are indicated.

Find authenticated court documents without watermarks at docketalarm.com.

522 ROSE AND GALLIONE

complementary to all but nine nucleotides at the 5' end of the G mRNA sequence (24). Similarly, the sequence of the pM309 insert (Fig. 2) showed that it contained all but 19 nucleotides from the 5' M mRNA sequence (24). Because the translation initiation codons for the G and M mRNA's are located 30 and 42 nucleotides from the 5' ends of the mRNA's (23, 24), these clones clearly contain the sequences encoding the NH₂ termini of G and M proteins. Subsequent sequence analysis showed that these clones contained the 3'-terminal mRNA sequences as well.



FIG. 2. Sequencing gels showing the sequences in the regions of pG1 and pM309 that correspond to the 5'-proximal regions of G and M mRNA's. Panels: A, sequence from the 5'-proximal AluI site of the pG1insert through the C-tails. Panel B, sequence from the pBR322 HpaII site proximal to the 5' end of pM309 reading into the insert DNA. Numbers indicate the alignment of the sequences with the mRNA sequence (see Fig. 4 and 6).

The mechanism of formation of pG1 and pM309 can be explained from the 5'-terminal sequences of the mRNAs (Fig. 4 and 6 of reference 24). The priming of second-strand DNA synthesis in pG1 presumably occurred from the 3' end of the cDNA looped back to base pair perfectly with positions 13 through 18, leaving an unpaired loop of six nucleotides. After second strand synthesis, digestion of the loop with nuclease S1 must have left the unpaired nucleotides 10 through 12 intact because they appear in the clone. In the formation of pM309, the 3' end of the cDNA presumably looped back to pair with positions 21 through 24 and began priming. Nuclease S1 digestion was complete or nearly complete, generating a clone lacking 19 nucleotides of the 5' mRNA sequence.

Sequencing strategy for pG1. Initial restriction mapping combined with previous partial sequences (24) allowed us to locate the two HaeIII sites within the pG1 insert (Fig. 3). Initial sequences were then established from the central HaeIII site reading 520 nucleotides toward the 5' end of the mRNA and 350 nucleotides toward the 3' end. Additional sequence was determined from the 3' HaeIII site for 520 nucleotides toward the 5' end of the mRNA. This approach, with sequencing gels 152 cm long, allowed us to establish a tentative sequence for almost the entire G gene using only two restriction sites. A restriction map was then established from this sequence by computer analysis (32). Subsequently we were able to identify specific restriction fragments required to complete the sequence on both DNA strands and correct a few errors that were generated by reading sequences more than 450 nucleotides from the labeled end. Figure 3 shows the locations of sites for restriction enzymes that were used in the sequence analysis and the specific regions sequenced. Several EcoRII sites (CC $\frac{A}{T}$ GG) were

encountered in this analysis. These sequences showed gaps in the sequence ladder at the second C residue (20), and the sequences at these sites were always confirmed by sequencing the complementary strand.

The complete sequence of the G mRNA predicted from the sequence of the pG1 clone (and previous data, references 24 and 27) is shown in Fig. 4. The sequence contains a single open reading frame for translation beginning at the 5'-proximal ATG codon (positions 29 through 31) and ending 100 nucleotides from the 3' end of the mRNA. That this is the reading frame encoding G protein has been confirmed by direct sequencing of G protein at both ends of the molecule (12, 15, 27).

Nucleotide sequencing of the M mRNA. A

J. VIROL.

Vol. 39, 1981



FIG. 3. Restriction map of the pGF1 insert DNA. Sites for restriction enzymes that were used in sequencing are shown. The left end corresponds to the 5' end of G mRNA. Arrows indicate the regions sequenced and the direction of sequencing. Numbers are in hundreds of nucleotides starting from the 5' end of the mRNA. Terminal PstI sites are at the junctions of pBR322 sequences with the dG:dC tails. The 5' terminal sequence of G mRNA determined previously from the genomic sequence is indicated (27). The sequence of the 3' terminal 470 nucleotides were reported previously from a smaller clone, but also were sequenced on at least one DNA strand in pG1.

sequence from the 5' end of the M protein mRNA had been determined from the sequence of a DNA primer extended from the adjacent NS gene along the genome into the M mRNA-coding region (24). This sequence overlaps and agrees with the sequences determined from the pM309 insert (Fig. 5) and with the sequence of a partial M clone (pM101; L. Iverson, unpublished results; 12a). The remainder of the M mRNA sequence was determined by sequencing from the restriction sites indicated in Fig. 5. All regions were sequenced on both DNA strands or sequenced at least twice (from different sites) to ensure accuracy.

The M mRNA is 831 nucleotides long without polyadenylic acid and contains a single open reading frame for translation extending from the 5'-proximal ATG codon at positions 42 through 44 to a TAG termination codon at positions 729 through 831 (Fig. 6). Other reading frames are blocked extensively throughout the mRNA sequence, leaving little doubt that the predicted M protein sequence is correct. Although no direct protein sequence is available for M, the sequence of the ribosome binding site determined directly from the mRNA (23) is in complete agreement with the sequence predicted here from the cDNA clone. The sequence also matches the partial RNA sequence of the M mRNA determined by priming on genomic RNA from the adjacent NS gene (24).

DISCUSSION

VSV G protein structure and modification. The nucleotide sequence presented here for the VSV glycoprotein mRNA predicts a protein of 511 amino acids (57,416 daltons), including the NH₂-terminal signal sequence of 16 amino acids that is not present on the mature protein (12, 15, 27). The predicted COOH-terminal sequence of G contains a hydrophobic domain of 20 amino acids (residues 463 through 482) followed by a hydrophilic domain (residues 483 through 511). We have presented evidence previously that the hydrophobic domain spans the lipid bilayer of the viral envelope, and that the COOH-terminal charged domain residues inside the viral envelope leaving more than 95% of the protein protruding from the virus (27). The COOH-terminal basic domain probably interacts with internal virion proteins and cellular proteins during budding of the virus from the plasma membrane. This portion of G could also play a role in targeting the protein to the plasma membrane.

The VSV G protein contains two apparently identical Asn-linked complex oligosaccharides (10, 21). The precise locations of the oligosaccharide attachment sites in G are of special interest because the timing of the glycosylation events relative to the timing of G protein synthesis has been examined in detail in vitro. Addition of oligosaccharide chains by transfer from a lipid carrier occurs on the nascent polypeptide chain (29, 34). The first addition occurs when about 38% of the chain has been synthesized, and the second occurs when about 70% has been synthesized (29), assuming that the rate of protein synthesis is uniform in the in vitro system. N-glycosidic linkage of oligosaccharides to proteins occurs at Asn-X-Ser or Asn-X-Thr sequences (19). Inspection of the predicted G protein sequence reveals 18 Asn residues, but only two of these (amino acid residues 178 and 335, Fig. 4) occur in canonical glycosylation sequences. We therefore presume that these are the actual glycosylation sites. These sites are at fractional distances of 0.35 and 0.66 from the NH₂ terminus. The nearly exact correspondence of the positions of these sites with fraction of G synthesized when glycosylation occurs suggests that transfer of oligosaccharides to the nascent protein chain occurs when the appropriate Asn residues traverse the rough endoplasmic reticu-

DOCKET



Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time** alerts and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

