JOURNAL OF VIROLOGY, Aug. 1984, p. 505–514
0022-538X/84/080505-10$02.00/0

Vol. 51, No. 2

Primary Structure of the Vesicular Stomatitis Virus Polymerase (*L*) Gene: Evidence for a High Frequency of Mutations

MANFRED SCHUBERT,* GEORGE G. HARMISON, AND ELLEN MEIER

Laboratory of Molecular Genetics, National Institute of Neurological and Communicative Disorders and Stroke, Bethesda, Maryland 20205

Received 5 December 1983/Accepted 10 February 1984

A consensus sequence of the polymerase (*L*) gene of vesicular stomatitis virus, derived from three genomic cDNA copies, is presented. This analysis completes the primary structure of the vesicular stomatitis virus genome, totaling 11,162 bases. The *L* gene alone spans 6,380 nucleotides and codes for a basic 2,109-amino-acid protein with a molecular weight of 241,012. Sixteen point mutations were detected among cDNA clones prepared from viral RNA of the same strain, representing direct evidence for either the high mutability of vesicular stomatitis virus, the infidelity of reverse transcription during cDNA synthesis, or a combination of both. Some mutations, if present in the viral genome, would result in the translation of incomplete L proteins. For example, two out of four cDNA copies which covered the same region of the *L* gene had a single-base deletion in the exact same postion, whereas the other two clones did not, strongly suggesting that a subpopulation of the genomic RNA may contain this lethal mutation. These lethal mutants define a new class of defective and most likely interfering particles which are indistinguishable in size from the parental virus and can be distinguished only by direct sequencing. We suggest that because of its infidelity, the viral polymerase itself introduces mutations and because of its size, most of these mutations are localized within the polymerase gene. In persistently infected cells in which the selective pressures on the polymerase are different, some of these *L* gene mutations may further erode the accuracy of the polymerase and thereby lead to the increased mutation rate that is characteristic of this type of infection.

The synthesis of functional mRNAs of nonsegmented negative-strand RNA viruses such as measles, Sendai, respiratory syncytial, Newcastle disease, and rabies viruses, vesicular stomatitis virus (VSV), and others requires an impressive number of diverse activities. These functions include specific binding of the polymerase to the ribonucleoprotein template, initiation, polymerization, capping, two types of methylation, polyadenylation, and specific termination (2, 4). In addition, the transcriptase recognizes highly conserved signal sequences along the template (26, 33, 44, 47). These signals are ignored during replication, a process which is dependent on protein synthesis (23) and possibly involves the cooperation of host factors.

In vitro transcription experiments with VSV demonstrated that besides the nucleocapsid template, only the viral NS and L proteins are necessary for transcription (11), suggesting that these proteins are capable of all of the enzymatic activities listed above. Since the NS and L proteins are individually inactive, they are considered dissimilar subunits of the polymerase complex. None of the activities mentioned above can be separated from the complex, and exogenous RNAs are not accepted as substrates for any of the modification functions such as capping, methylation, and polyadenylation. It appears not only that all of these diverse activities are carried out on the template, but also that they require nascent transcripts as substrates.

Many nonsegmented negative-strand virus genomes code for a large protein (L) with a molecular weight of approximately 200,000 and a phosphoprotein (P) (NS of VSV) with a molecular weight of approximately 25,000. The relative positions of these two genes within the genomes of nonsegmented negative-strand viruses are identical (3′-*N*,*P* . . . *L*-5′) (1, 3). In those cases examined to date, the gene for the small phosphoprotein P is always positioned next to the nucleocapsid protein gene. The VSV *NS* mRNA is the second most abundant in infected cells. The least abundant transcription product is derived from the *L* gene, which often covers the entire 5′ half of the genome. Because of its giant size (241,000 for VSV; see below), we anticipate that most of the activities of the multifunctional polymerase complex are performed by the L protein.

In this communication, we report the first nucleotide sequence of an *L* gene of a negative-strand virus. This single gene of VSV consists of 6,380 nucleotides and is about the size of the entire tobacco mosaic virus genome (6,395 bases). Together with the reported sequences of the other VSV genes *N*, *NS*, *M*, and *G* (16, 45), this analysis completes the primary structure of the VSV genome. The L protein alone covers more than 60% of the total coding regions of VSV mRNAs.

There is no evidence for proofreading mechanisms during genome replication, which may explain why RNA viruses exhibit a very high mutation rate, contributing to the rapid evolution and variability of these viruses (21). Within the *L* gene, at least 20 nucleotide differences between various cDNA clones of the same regions were detected. These differences, if expressed, would appear not only as amino acid substitutions in the L protein, but surprisingly also in the synthesis of incomplete L proteins owing to the creation of stop codons in the reading frame. The importance of these findings for the biology of the virus as well as for the determination of nucleotide sequences of RNA viruses in general is discussed.

MATERIALS AND METHODS

Virus growth and RNA purification. The VSV Mudd Summers strain of the Indiana serotype was used for the

* Corresponding author.

isolation of cDNA copies. The virus was plaque purified, and this virus stock was used after an additional passage to inoculate BHK cells at a multiplicity of infection of 10 [1]. The virus particles were banded on sucrose gradients. The three cDNA libraries (see below) were made from RNA which was isolated from virus particles derived from the same stock. Propagation of the virus and the purification of the RNAs on sucrose gradients have been described earlier (28, 31, 35). The defective interfering (DI) particle RNAs which were used as hybridization probes were derived either from the parental Mudd Summers strain (DI 011 and 611) or from the San Juan strain [DI T and T(L)]. DI T(L) was a generous gift from Suzanne U. Emerson, University of Virginia, Charlottesville. The RNAs of these DI particles have been characterized previously (30) and have now been precisely mapped by using the nucleotide sequence of the *L* gene described in this paper (34a).

Preparation of cDNA clones. The cDNA clones 011-2 and HR3-9 were derived by copying the DI particle RNA of DI 011 and the heat-resistant strain (HR) of the Mudd Summers isolate, respectively. These cDNA clones have been described earlier and were partially sequenced (56). The other 30 clones were members of three cDNA libraries designated MS 4, 21-4, 23-1, 26, 38, 114, 160, 177, 186, 216, 218, 282, 396, 398, and 437; JS 18, 24, 36, 39, 40, 43, 58-3, 60, and 67; and LH 321, 455, 602, 649, 661, and 679. The reverse transcription reactions were carried out essentially as described previously (49), except that with the MS clones the first strand was oligodeoxyadenylic acid tailed, and the second strand was specifically primed by using a short oligodeoxythymidylic acid primer (29). All reverse transcriptase reactions contained 10 U of RNasin (Biotec, Madison, Wis.). After nuclease S1 treatment, the double-stranded DNA was oligodeoxycytidylic acid tailed and cloned into the oligodeoxyguanidylic acid-tailed *Pst*I site of plasmid pBR322 according to standard procedures (9, 53). *Escherichia coli* HB101 cells were transformed, and the tetracycline-resistant, ampicillin-sensitive colonies were picked.

Identification of the clones. 42S genomic RNA as well as the RNAs of DI 011, T, T(L), and 611 were partially hydrolyzed in 50 mM Na_2CO_3/$NaHCO_3$–1 mM EDTA, pH 9 (10), for 8 min at 90°C. The fragments, averaging about 150 nucleotides in length, were labeled at their 5′ ends with [γ-^{32}P]ATP (3,000 Ci/mmol) and polynucleotide kinase (42). The labeled fragments were used during colony hybridizations as previously described (18). In addition, 3′-labeled cDNA plasmid inserts of the *L* gene region were used as hybridization probes.

Nucleotide sequence analysis. Restriction fragments were either labeled at their 3′ termini with [α-^{32}P]cordycepin triphosphate and terminal deoxynucleotidyl transferase (50) or labeled after phosphatase treatment at their 5′ ends with [γ-^{32}P]ATP and polynucleotide kinase (42). After secondary restriction cuts, the fragments were separated on agarose gels and were bound to and eluted from short DEAE membrane strips. The fragments were then sequenced by the chemical sequencing procedure of Maxam and Gilbert (32).

Computer analysis of the sequences. The search for overlapping sequences among the partial sequences of the clones, restriction site analysis, the translation of the message sequence, and the determination of the amino acid composition were done with an IBM 370 computer with the programs of Queen and Korn (40). Comparison of the sequences with those in the Genetic Sequence Data Bank in Los Alamos, N.M., and in the Protein Sequence Data Base of the National Biomedical Research Foundation in Washington, D.C., were performed with an algorithm developed by Wilbur and Lipman (54).

RESULTS

Cloning of the *L* gene. As a first step in dissecting the individual functions of the polymerase complex, we cloned and sequenced the *L* gene. 42S genomic RNA of VSV, Indiana serotype, was reverse transcribed by random self-priming. After second-strand synthesis and oligodeoxycytidylic acid tailing, the DNA fragments were inserted into the deoxyguanidylic acid-tailed *Pst*I restriction site of the bacterial plasmid pBR322. *E. coli* HB101 cells were transformed, and more than 1,200 tetracycline-resistant and ampicillin-sensitive colonies were picked and analyzed by filter hybridization. 5′-^{32}P-end-labeled RNA fragments of DI particle RNAs such as DI 011, T, T(L), and 611 as well as end-labeled fragments of the total 42S genome were used as hybridization probes. The DI RNAs used lack sequences from the 3′ half of the genome and represent overlapping regions of the 5′ end of the genome, the *L* gene region (30).

More than 500 cDNA clones ranging from 100 to 3,400 base pairs were tentatively mapped within the *L* gene region. Clones from the 3′ end of the *L* gene were identified by using a cDNA clone from the heat-resistant strain of VSV Indiana which spans the *G* and *L* boundary and extends into the *L* gene for 1,200 base pairs (56). The proportion of cDNA clones which were derived from the *L* gene corresponded roughly to the ratio of the size of the *L* gene relative to the complete length of the VSV genome, indicating that priming by the reverse transcriptase occurred randomly along the genomic RNA, despite the absence of random oligonucleotide primers. Sequence analysis of some, but not all, cDNAs revealed that a few cDNA clones which covered long stretches of the *L* gene region also contained at their 3′ ends short stretches of sequences derived from, e.g., the *NS* gene of VSV, a region more than 9,000 bases upstream. We suspect that short RNA molecules which were copurified with the genomic RNA template or breakdown products of the template itself served as primers during the reverse transcriptase reaction.

Nucleotide sequence analysis. The map of 32 cDNA clones within the *L* gene which were partially or completely used to determine its nucleotide sequence is shown in Fig. 1. The 5′ termini of the VSV genome and the *L* gene as well as the *G* and *L* gene boundary are indicated. DNA fragments were labeled either at their 5′ ends with [γ-^{32}P]ATP and polynucleotide kinase or at their 3′ ends with [α-^{32}P]cordycepin triphosphate and terminal deoxynucleotidyl transferase. After secondary cuts with appropriate restriction endonucleases, the fragments were isolated from agarose gels and subjected to the chemical sequencing method described by Maxam and Gilbert (32). The positions of the partial sequences of either (−) or (+) polarity are indicated above and below the map of the cDNA clones (Fig. 1). The total number of nucleotides sequenced was more than 32,000, derived from approximately 19,000 nucleotides of 32 independent clones. Thus, each nucleotide of the *L* gene was, on the average, sequenced five times in (+) and (−) sense by using three independently derived cDNA copies. The sequence in Fig. 2, therefore, represents a consensus sequence of, on the average, three cDNA copies. This extensive sequence analysis was necessary, as will be outlined below.

Translation of the *L* message. Six complementation groups of conditional mutants have been identified with the New Jersey serotype of VSV (39), but only five viral proteins have been reported. The massive size of the *L* gene region

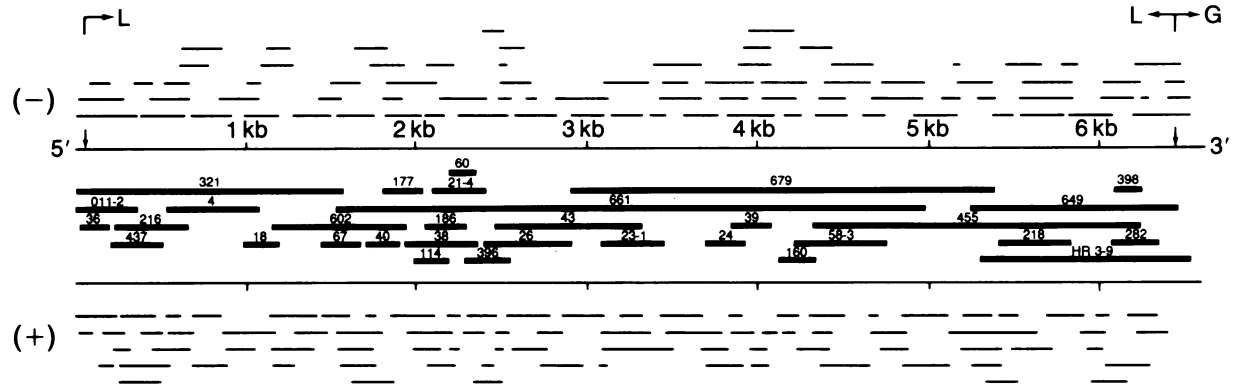

FIG. 1. Map of cDNA clones within the *L* gene. The 32 cDNA clones listed (black bars) were either partially or completely used to determine the nucleotide sequence of the *L* gene. The 5′ terminus corresponds to the 5′ end of the VSV genome. The *G* and *L* boundary and the polyadenylation site of the *L* message near the 5′ terminus are indicated by arrows. The stretches of partial sequences in (−) or (+) polarity were combined to the consensus sequence of the *L* message shown in Fig. 2.

raised the possibility for the coding potential of a sixth viral peptide. Figure 3 shows the translation of the *L* message as well as VSV genomic RNA in three reading frames. Regions which were at least 70 nucleotides long and did not contain a translational stop codon are boxed. The beginning and end of each box mark termination codons. Closed boxes indicate the presence of at least one AUG start codon within that region, independent of its exact position. It is obvious that reading frame 2 of the *L* message (Fig. 3, A) belongs to the giant L protein. VSV transcribes and replicates in the cytoplasm of the host cell, and there is no evidence of splicing. Our nucleotide sequence analysis clearly rules out the presence of overlapping reading frames, strongly suggesting that only one protein is encoded by the *L* message. The second largest open reading frame was found in genomic RNA (Fig. 3, B), which contains a start codon and has a coding potential for a peptide with a molecular weight of about 10,000. However, there is no evidence for genomic sense-translatable transcripts or for the existence of substantial amounts of genomic sense RNA as free RNA in the cytoplasm.

The length of the *L* message is 6,380 nucleotides, including the first seven A residues of the polyadenylic acid [poly(A)] tail. The ribosome binding site of the *L* message, which includes the first AUG in position 11 of the message, has been reported earlier (43). We have previously isolated and described the polyadenylation site of the *L* message located at positions 60 to 66 from the 5′ end of the genome (47, 48). Therefore, the complete amino acid sequence of the L protein can be deduced from the message sequence (Fig. 2). Translation of the L protein starts with methionine at position 11 of the message and terminates with aspartic acid before the UAA translational stop 43 nucleotides before the poly(A) tail. Despite its size, the *L* message contains the smallest amount of nontranslated terminal sequences among all VSV messages, 53 bases. It contains a single open reading frame for a 2,109-amino-acid protein with a calculated molecular weight of 241,012, exceeding earlier size determinations. The summary of the codon usage for the translation of the L protein revealed that, as with the other VSV proteins, codons containing CG are used less frequently during translation, a feature also observed during influenza virus translation, affecting the codon usage for serine, proline, threonine, alanine, and particularly arginine. This underrepresentation may reflect the adaptation of these viruses to the availability of the corresponding host tRNAs.

The L protein. The amino acid composition of the L protein is given in Table 1. The amounts of basic, acidic, polar, and nonpolar amino acids are summarized. The L protein is a basic protein, but less basic than the VSV M protein, which contains 17.1% basic and 11.4% acidic amino acids. Unlike the NS, M, and G proteins (16, 45), L does not exhibit relatively large clusters of amino acids of similar characteristics, e.g., charged, hydrophobic. Although the overall amounts of basic, acidic, polar, and nonpolar amino acids are almost identical to those in the PB2 protein of influenza virus (13), the *trans*-capping protein (51), the relative amounts of the individual amino acids are quite different. Computer comparison of the nucleotide and amino acid sequences of the L and the three polymerase proteins of influenza virus (PB1, PB2, and PA) (13, 55), which combined are roughly the size of the *L* gene protein, did not reveal any significant homology. In fact, comparison of the nucleotide and amino acid sequences of the L protein and all of those stored in the Genetic Sequence Data Bank in Los Alamos and the Protein Sequence Data Base of the National Biomedical Research Foundation in Washington, D.C., did not reveal any significant homology to any of the sequences, including those of a number of proteins which interact with nucleic acids. The total number of bases screened was 2×10^6. A comparison with the L proteins of other negative-strand viruses awaits nucleotide sequence information from these viruses.

Mutations within the *L* gene. As outlined above, each nucleotide of the *L* gene was sequenced, on the average, five times in the plus and minus senses with three independently derived cDNA copies. This extensive nucleotide sequence analysis was necessary because at least 20 base differences between various cDNA clones from the same regions were detected. We would like to emphasize that we cannot distinguish at this time whether all of the differences are actually present in subpopulations of VSV genomic RNA or whether they are simply errors of reverse transcription (see below). A map of these 20 mutational changes within the *L* message is shown in Fig. 4. The identification numbers of the cDNA clones which differ from the consensus sequence are indicated. Most of the changes included point mutations which, if present in the viral genome, would lead to the

```
AACAGCAAUCAUGGAAGUCCACGAUUUUGAGACCGACGAGUUCAAUGAUUUCAAUGAAGAUGACUAUGCCACAAGAGAAUUCCUGAAUCCCGAUGAGCGC    100
                     METGluValHisAspPheGluThrAspGluPheAsnAspPheAsnGluAspAspTyrAlaThrArgGluPheLeuAsnProAspGluArg  (  30)

AUGACGUACUUGAAUCAUGCUGAUUACAAUUUGAAUUCUCCUCUAAUUAGUGAUGAUAUUGACAAUUUGAUCAGGAAAUUCAAUUCUCUUCCGAUUCCCUCGAUGUGGGAUAGUAAGAAC    220
METThrTyrLeuAsnHisAlaAspTyrAsnLeuAsnSerProLeuIleSerAspAspIleAspAsnLeuIleArgLysPheAsnSerLeuProIleProSerMETTrpAspSerLysAsn  (  70)

UGGGAUGGAGUUCUUGAGAUGUUAACAUCAUGUCAAGCCAAUCCCAUCUCAACAUCUCAGAUGCAUAAAUGGAUGGGAAGUUGGUUAAUGUCUGAUAAUCAUGAUGCCAGUCAAGGGUAU    340
TrpAspGlyValLeuGluMETLeuThrSerCysGlnAlaAsnProIleSerThrSerGlnMETHisLysTrpMETGlySerTrpLeuMETSerAspAsnHisAspAlaSerGlnGlyTyr  ( 110)

AGUUUUUUACAUGAAGUGGACAAAGAGGCAGAAAUAACAUUUGACGUGGUGGAGACCUUCAUCCGCGGCUGGGGCAACAAACCAAUUGAAUACAUCAAAAAGGAAAGAUGGACUGACUCA    460
SerPheLeuHisGluValAspLysGluAlaGluIleThrPheAspValValGluThrPheIleArgGlyTrpGlyAsnLysProIleGluTyrIleLysLysGluArgTrpThrAspSer  ( 150)

UUCAAAAUUCUCGCUUAUUUGUGUCAAAAGUUUUUGGACUUACACAAGUUGACAUUAAUCUUAAAUGCUGUCUCUGAGGUGGAAUUGCUCAACUUGGCGAGGACUUUCAAAGGCAAAGUC    580
PheLysIleLeuAlaTyrLeuCysGlnLysPheLeuAspLeuHisLysLeuThrLeuIleLeuAsnAlaValSerGluValGluLeuLeuAsnLeuAlaArgThrPheLysGlyLysVal  ( 190)

AGAAGAAGUUCUCAUGGAACGAACAUAUGCAGGAUUAGGGUUCCCAGCUUGGGUCCUACUUUUAUUUCAGAAGGAUGGGCUUACUUCAAGAAACUUGAUAUUCUAAUGGACCGAAACUUU    700
ArgArgSerSerHisGlyThrAsnIleCysArgIleArgValProSerLeuGlyProThrPheIleSerGluGlyTrpAlaTyrPheLysLysLeuAspIleLeuMETAspArgAsnPhe  ( 230)

CUGUUAAUGGUCAAAGAUGUGAUUAUAGGGAGGAUGCAAACGGUGCUAUCCAUGGUAUGUAGAAUAGACAACCUGUUCUCAGAGCAAGACAUCUUCUCCCUUCUAAAUAUCUACAGAAUU    820
LeuLeuMETValLysAspValIleIleGlyArgMETGlnThrValLeuSerMETValCysArgIleAspAsnLeuPheSerGluGlnAspIlePheSerLeuLeuAsnIleTyrArgIle  ( 270)

GGAGAUAAAAUUGUGGAGAGGCAGGGAAAUUUUUCUUAUGACUUGAUUAAAAUGGUGGAACCGAUAUGCAACUUGAAGCUGAUGAAAUUAGCAAGAGAAUCAAGGCCUUUAGUCCCACAA    940
GlyAspLysIleValGluArgGlnGlyAsnPheSerTyrAspLeuIleLysMETValGluProIleCysAsnLeuLysLeuMETLysLeuAlaArgGluSerArgProLeuValProGln  ( 310)

UUCCCUCAUUUUGAAAAUCAUAUCAAGACUUCUGUUGAUGAAGGGGCAAAAAUUGACCGAGGUAUAAGAUUCCUCCAUGAUCAGAUAAUGAGUGUGAAAACAGUGGAUCUCACACUGGUG    1060
PheProHisPheGluAsnHisIleLysThrSerValAspGluGlyAlaLysIleAspArgGlyIleArgPheLeuHisAspGlnIleMETSerValLysThrValAspLeuThrLeuVal  ( 350)

AUUUAUGGAUCGUUCAGACAUUGGGGUCAUCCUUUUAUAGAUUAUUACACUGGACUAGAAAAAUUACAUUCCCAAGUAACCAUGAAGAAAGAUAUUGAUGUGUCAUAUGCAAAAGCACUU    1180
IleTyrGlySerPheArgHisTrpGlyHisProPheIleAspTyrTyrThrGlyLeuGluLysLeuHisSerGlnValThrMETLysLysAspIleAspValSerTyrAlaLysAlaLeu  ( 390)

GCAAGUGAUUUAGCUCGGAUUGUUCUAUUUCAACAGUUCAAUGAUCAUAAAAAGUGGUUCGUGAAUGGAGACUUGCUCCCUCAUGAUCAUCCCUUUAAAAGUCAUGUUAAAGAAAAUACA    1300
AlaSerAspLeuAlaArgIleValLeuPheGlnGlnPheAsnAspHisLysLysTrpPheValAsnGlyAspLeuLeuProHisAspHisProPheLysSerHisValLysGluAsnThr  ( 430)

UGGCCCACAGCUGCUCAAGUUCAAGAUUUUGGAGAUAAAUGGCAUGAACUUCCGCUGAUUAAAUGUUUUGAAAUACCCGACUUACUAGACCCAUCGAUAAUAUACUCUCACAAAAGUCAU    1420
TrpProThrAlaAlaGlnValGlnAspPheGlyAspLysTrpHisGluLeuProLeuIleLysCysPheGluIleProAspLeuLeuAspProSerIleIleTyrSerAspLysSerHis  ( 470)

UCAAUGAAUAGGUCAGAGGUGUUGAAACAUGUCCGAAUGAAUCCGAACACUCCUAUCCCUAGUAAAAAGGUGUUGCAGACUAUGUUGGACACAAAGGCUACCAAUUGGAAAGAAUUUCUU    1540
SerMETAsnArgSerGluValLeuLysHisValArgMETAsnProAsnThrProIleProSerLysLysValLeuGlnThrMETLeuAspThrLysAlaThrAsnTrpLysGluPheLeu  ( 510)

AAAGAGAUUGAUGAGAAGGGCUUAGAUGAUGAUGAUCUAAUUAUUGGUCUUAAAGGAAAGGAGAGGGAACUGAAGUUGGCAGGUAGAUUUUUCUCCCUAAUGUCUUGGAAAUUCCGAGAA    1660
LysGluIleAspGluLysGlyLeuAspAspAspAspLeuIleIleGlyLeuLysGlyLysGluArgGluLeuLysLeuAlaGlyArgPhePheSerLeuMETSerTrpLysLeuArgGlu  ( 550)

UACUUUGUAAUUACCGAAUAUUUGAUAAAGACUCAUUUCGUCCCUAUGUUUAAAGGCCUGACAAUGGCGGACGAUCUAACUGCAGUCAUUAAAAAGAUGUUAGAUUCCUCAUCCGGCCAA    1780
TyrPheValIleThrGluTyrLeuIleLysThrHisPheValProMETPheLysGlyLeuThrMETAlaAspAspLeuThrAlaValIleLysLysMETLeuAspSerSerSerGlyGln  ( 590)

GGAUUGAAGUCAUAUGAGGCAAUUUGCAUAGCCAAUCACAUUGAUUACGAAAAAUGGAAUAACCACCAAAGGAAGUUAUCAAACGGCCCAGUGUUCCGAGUUAUGGGCCAGUUCUUAGGU    1900
GlyLeuLysSerTyrGluAlaIleCysIleAlaAsnHisIleAspTyrGluLysTrpAsnAsnHisGlnArgLysLeuSerAsnGlyProValPheArgValMETGlyGlnPheLeuGly  ( 630)

UAUCCAUCCUUAAUCGAGAGAACUCAUGAAUUUUUUGAGAAAAGUCUUAUAUACUACAAUGGAAGACCAGACUUGAUGCGUGUUCACAACAACACACUGAUCAAUUCAACCUCCCAACGA    2020
TyrProSerLeuIleGluArgThrHisGluPhePheGluLysSerLeuIleTyrTyrAsnGlyArgProAspLeuMETArgValHisAsnAsnThrLeuIleAsnSerThrSerGlnArg  ( 670)

GUUUGUUGGCAAGGACAAGAGGGUGGACUGGAAGGUCUACGGCAAAAAGGAUGGACUAUCCUCAAUCUACUGGUUAUUCAAAGAGAGGCUAAAAUCAGAAACACUGCUGUCAAAGUCUUG    2140
ValCysTrpGlnGlyGlnGluGlyGlyLeuGluGlyLeuArgGlnLysGlyTrpSerIleLeuAsnLeuLeuValIleGlnArgGluAlaLysIleArgAsnThrAlaValLysValLeu  ( 710)

GCACAAGGUGAUAAUCAAGUUAUUUGCACACAGUAUAAAACGAAGAAAUCGAGAAACGUUGUAGAAUUACAGGGUGCUCUCAAUCAAAUGGUUUCUAAUAAUGAGAAAAUUAUGACUGCA    2260
AlaGlnGlyAspAsnGlnValIleCysThrGlnTyrLysThrLysLysSerArgAsnValValGluLeuGlnGlyAlaLeuAsnGlnMETValSerAsnAsnGluLysIleMETThrAla  ( 750)

AUCAAAAUAGGGACAGGGAAGUUAGGACUUUUGAUAAAUGACGAUGAGACUAUGCAAUCUGCAGAUUACUUGAAUUAUGGAAAAAUACCGAUUUUCCGUGGAGUGAUUAGAGGGUUAGAG    2380
IleLysIleGlyThrGlyLysLeuGlyLeuLeuIleAsnAspAspGluThrMETGlnSerAlaAspTyrLeuAsnTyrGlyLysIleProIlePheArgGlyValIleArgGlyLeuGlu  ( 790)

ACCAAGAGAUGGUCACGAGUGACUUGUGUCACCAAUGACCAAAUACCCACUUGUGCUAAUAUAAUGAGCUCAGUUUCCACAAAUGCUCUCACCGUAGCUCAUUUUGCUGAGAACCCAAUC    2500
ThrLysArgTrpSerArgValThrCysValThrAsnAspGlnIleProThrCysAlaAsnIleMETSerSerValSerThrAsnAlaLeuThrValAlaHisPheAlaGluAsnProIle  ( 830)

AAUGCCAUGAUACAGUACAAUUAUUUUGGGACAUUUGCUAGACUCUUGUUGAUGAUGCAUGAUCCUGCUCUUCGUCAAUCAUUGUAUGAAGUUCAAGAUAAGAUACCGGGCUUGCACAGU    2620
AsnAlaMETIleGlnTyrAsnTyrPheGlyThrPheAlaArgLeuLeuLeuMETMETHisAspProAlaLeuArgGlnSerLeuTyrGluValGlnAspLysIleProGlyLeuHisSer  ( 870)

UCUACUUUCAAAUACGCCAUGUUGUAUUUGGACCCUUCCAUUGGAGGAGUGUCGGGCAUGUCUUUGUCCAGGUUUUUGAUUAGAGCCUUCCCAGAUCCCGUAACAGAAAGUCUCUCAUUC    2740
SerThrPheLysTyrAlaMETLeuTyrLeuAspProSerIleGlyGlyValSerGlyMETSerLeuSerArgPheLeuIleArgAlaPheProAspProValThrGluSerLeuSerPhe  ( 910)

UGGAGAUUCAUCCAUGUACAUGCUCGAAGUGAGCAUCUGAAGGAGAUGAGUGCAGUAUUUGGAAACCCCGAGAUAGCCAAGUUUCGAAUAACUCACAUAGACAAGCUAGUAGAAGAUCCA    2860
TrpArgPheIleHisValHisAlaArgSerGluHisLeuLysGluMETSerAlaValPheGlyAsnProGluIleAlaLysPheArgIleThrHisIleAspLysLeuValGluAspPro  ( 950)

ACCUCUCUGAACAUCGCUAUGGGAAUGAGUCCAGCGAACUUGUUAAAGACUGAGGUUAAAAAAUGCUUAAUCGAAUCAAGACAAACCAUCAGGAACCAGGUGAUUAAGGAUGCAACCAUA    2980
ThrSerLeuAsnIleAlaMETGlyMETSerProAlaAsnLeuLeuLysThrGluValLysLysCysLeuIleGluSerArgGlnThrIleArgAsnGlnValIleLysAspAlaThrIle  ( 990)

UAUUUGUAUCAUGAAGAGGAUCGGCUCAGAAGUUUCUUAUGGUCAAUAAAUCCUCUGUUCCCUAGAUUUUUAAGUGAAUUCAAAUCAGGCACUUUUUUGGGAGUCCCAGACGGGCUCAUC    3100
TyrLeuTyrHisGluGluAspArgLeuArgSerPheLeuTrpSerIleAsnProLeuPheProArgPheLeuSerGluPheLysSerGlyThrPheLeuGlyValAlaAspGlyLeuIle  (1030)

AGUCUAUUUCAAAAUUCUCGUACUAUUCGGAACUCCUUUAAGAAAAAGUAUCAUAGGGAAUUGGAUGAUUUGAUUGUGAGGAGUGAGGUAUCCUCUUUGACACAUUUAGGGAAACUUCAU    3220
SerLeuPheGlnAsnSerArgThrIleArgAsnSerPheLysLysLysTyrHisArgGluLeuAspAspLeuIleValArgSerGluValSerSerLeuThrHisLeuGlyLysLeuHis  (1070)

UUGAGAAGGGGAUCAUGUAAAAUGUGGACAUGUUCAGCUACUCAUGCUGACACAUUAAGAUACAAAUCCUGGGGCCGUACAGUUAUUGGGACAACUGUACCCCAUCCAUUAGAAAUGUUG    3340
LeuArgArgGlySerCysLysMETTrpThrCysSerAlaThrHisAlaAspThrLeuArgTyrLysSerTrpGlyArgThrValIleGlyThrThrValProHisProLeuGluMETLeu  (1110)

GGUCCACAACAUCGAAAAGAGACUCCUUGUGCACCAUGUAACACAUCAGGGUUCAAUUAUGUUUCUGUGCAUUGUCCAGACGGGAUCCAUGACGUCUUUAGUUCACGGGGACCAUUGCCU    3460
GlyProGlnHisArgLysGluThrProCysAlaProCysAsnThrSerGlyPheAsnTyrValSerValHisCysProAspGlyIleHisAspValPheSerSerArgGlyProLeuPro  (1150)
```

Fig. 2.—Continued

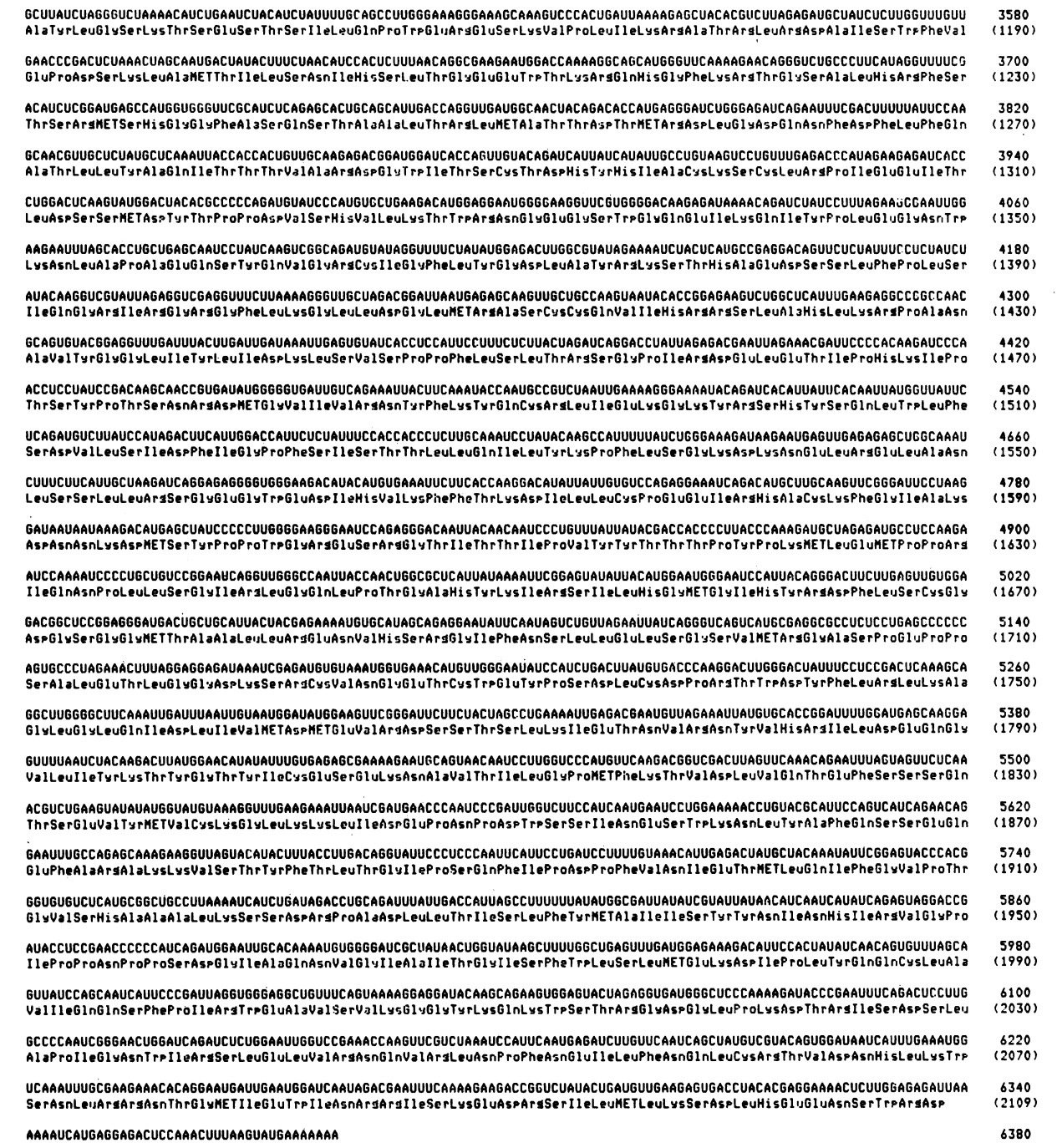

FIG. 2. Nucleotide sequence of *L* mRNA and the amino acid sequence of the L protein. Each nucleotide of the *L* message was, on the average, sequenced five times in the (+) and (−) senses. The total number of nucleotides determined was more than 32,000, derived from approximately 19,000 nucleotides of different cDNA copies. The sequence represents a consensus sequence of, on the average, three cDNA copies. The 5′-terminal methylated cap structure as well as the 3′-terminal poly(A) tail are not shown. The amino acid sequence of the L protein was deduced from the nucleotide sequence. The *L* message spans 6,380 nucleotides, including the first 7 A residues of the poly(A) tail, and codes for a basic 2,109-amino-acid long protein with a molecular weight of 241,012.

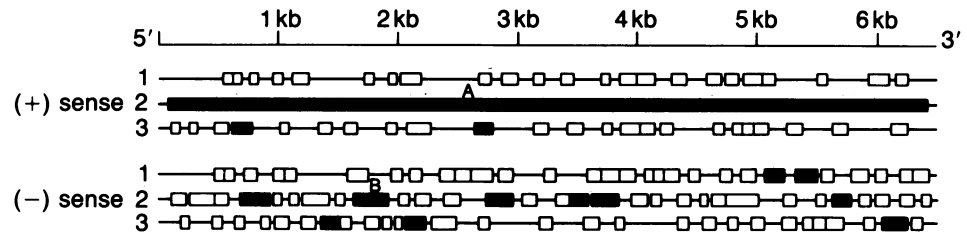

FIG. 3. Translation of the *L* message and *L* gene sequences. *L* mRNA and the (−) sense genomic RNA were both translated in three reading frames. Regions without a translational stop codon and at least 70 nucleotides in length are boxed. Closed boxes mark regions containing the translational start codon AUG. A represents the open reading frame for the L protein. B corresponds to the second largest open reading frame. It is unlikely that it is translated.

substitution of amino acids. In addition, single-base insertions and deletions were detected, which by shifting the reading frame would cause the synthesis of an incomplete L protein. The distribution of these mutations within the *L* message did not reveal any significant clustering of mutations in particular regions.

A summary of the mutations within the *L* gene and their potential effect on the L protein translation is shown in Table 2. At least two identical cDNA clones make up the consensus sequence and are defined as standard cDNA clones. These are compared with the altered cDNA clone. The positions of the mutations within the *L* message are indicated, and so are their effects on the primary structure of the L protein. The 15 base substitutions listed result in four silent mutations, without a change of the amino acid, and in amino acid changes ranging form Ile → Leu or Glu → Asp, without alterations in the amino acid characteristics, to significant changes such as Phe → Cys. This limited analysis does not reveal any preference for mutations in a particular position within the translation codon: the frequency at which changes were found was 5-4-6, corresponding to the positions 1, 2, and 3 of the codon.

The most interesting mutations in the sequence, however, consisted of single-base insertions and single-base deletions found in four different cDNA clones. These mutations, if expressed, would dramatically affect the primary structure of the L protein in that only portions of the protein ranging in size from about 27% (clone 455) to 92% (clone 216) could be translated, all resulting in an abortive translation of the L protein. We have isolated four cDNA clones which cover positon 4047 in the message (Table 2). Two cDNA clones (clones 21 and 396) showed exactly the same deletion, whereas the other two standard clones (clones 53 and 661) did not. The two cDNA clones 21 and 396, which contained the deletion, mapped at slightly different positions within the *L* gene (Fig. 1), which clearly rules out the possibility of repeated isolation of a single colony. In addition, clone 396 contained base substitutions in positions 4120 and 4054 that were not present in clone 21, whereas clone 21 contained three substitutions in positions 4062, 4063, and 4064, affecting two translation codons. The sequence of the corresponding region in clone 396 has not been determined. The fact that the identical deletion was present in the very same position in two different independently derived cDNA copies strongly suggests that a subpopulation of VSV genomic RNA contains this deletion and that the deletion is not the result of a random mistake during reverse transcription, but rather is an error by the viral polymerase during replication. If the reverse transcriptase made the mistake, it would have had to make the same mistake twice in the very same position. Deletion or insertion mutants of this type would define a new class of defective particles, which have the same size as the parental virus and can be identified only by direct sequencing. We suspect that these lethal mutants would interfere with the replication of the parental virus, like some conditional lethal mutants (*ts* mutants) of the *L* gene under the restrictive conditions (57).

Interestingly, all insertions or deletions involved U residues in the message sense, corresponding to A residues in the genomic sense. Additions of A residues which are not template specified have been reported in the genomic RNA of DI LT2 (25) and also during the transcription of leader RNA, which sometimes contains additional 3′-terminal A residues which are not encoded in the template. We have previously shown that the viral polymerase polyadenylates VSV messages by a chattering or slippage mechanism at a stretch of seven U residues on the template (19, 47, 48). The infidelity of this viral function may be involved in the insertion or deletion of U or A residues.

TABLE 1. Amino acid composition of the VSV L protein[a]

Amino acid	No. of residues
Ala	92
Arg	124
Asn	97
Asp	120
Cys	34
Gln	70
Glu	112
Gly	127
His	64
Ile	155
Leu	219
Lys	129
Met	60
Phe	93
Pro	98
Ser	178
Thr	122
Trp	42
Tyr	70
Val	103

[a] Total, 2, 109 amino acids. Basic, 15.0%; acidic, 11.0%; polar, 31.9%; nonpolar, 42.1%.

DISCUSSION

The VSV genome and its coding regions. The nucleotide sequence analysis of the *L* gene presented here, together with the sequences of the other four genes, *N*, *NS*, *M*, and *G* (16, 45), completes the primary structure of the VSV

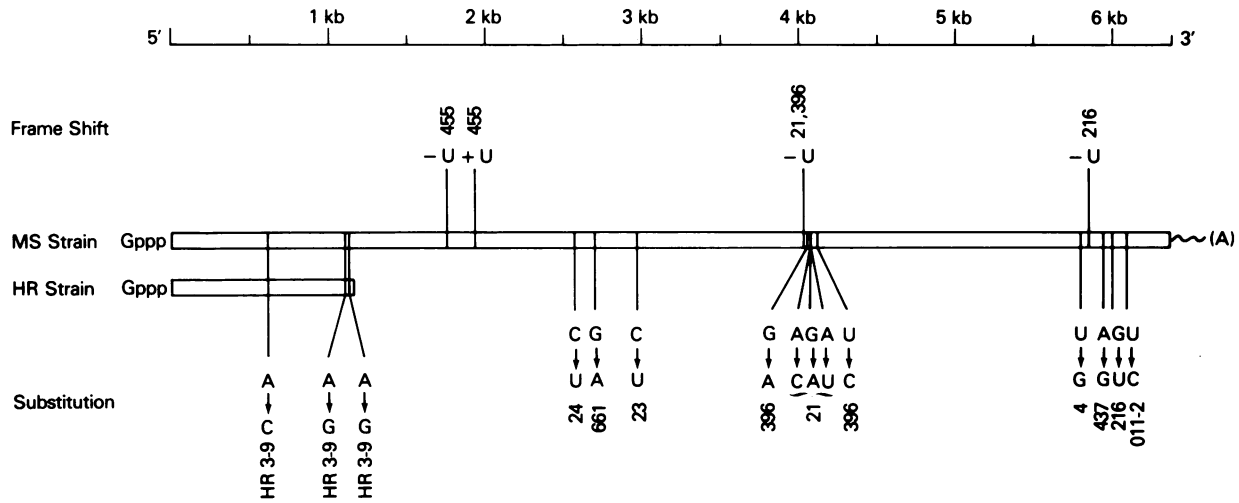

FIG. 4. Map of the mutations found within the *L* message. Mutations were not only detected between the Mudd Summers (MS) and the heat resistant (HR) strains of VSV, but also between various cDNA copies of genomes of the same strain (MS). Twenty nucleotide differences from the consensus sequence (Fig. 2) were found. Their positions within the *L* message are indicated, as well as the identification numbers of the cDNA clones.

genome. Figure 5 shows the positions and lengths of the five viral genes, including the intercistronic and terminal extracistronic regions. The genome, consisting of the leader region (l) (8, 27, 34, 46), the five genes, the four intercistronic regions (33, 44), and the 5′ terminal trailer region (t) (48), is 11,162 nucleotides long. Earlier size determinations of the genome which were based on complete nuclease digestions of genomic RNA predicted a total length of 11,278 nucleotides (41). The *L* gene alone makes up 57.2% of the genome, starting at position 4724 from the 3′ end of the genome. Almost the entire genome is transcribed (99.4%) into the short leader RNA and the five messages. The total coding region of the VSV genome is 93.9%, leaving 6.1% for the untranslated terminal message regions and the leader and the trailer regions. From this coding capacity, the L protein forms 60.4%, compared with 6.3% for the NS protein, which is approximately 1/10 the size of L.

Size and potential functions of the L protein. Since the

TABLE 2. VSV *L* gene mutations and their potential effects on the primary structure of the L protein[a]

Type of mutation	cDNA clone(s)		Position in *L* message	Sequence		Effect on L protein
	Standard	Altered		Standard	Altered	
Insertion	58-3, 679	455	1937	U_6	U_7	Incomplete
Deletion	321, 437	216	5853	GAG*U*AGG	GAGAGG	Incomplete
	53, 661	21	4047	CUU*U*AGA	CUUAGA	Incomplete
	53, 661	396	4047	CUU*U*AGA	CUUAGA	Incomplete
	58-3, 679	455	1750	CAU*U*AAA	CAUAAA	Incomplete
Substitution	216, 321	011-2	6087	A*U*U	A*C*U	Ile → Thr
	321, 437	216	6013	GA*G*	GA*U*	Glu → Asp
	216, 321	437	5942	*A*UG	*G*UG	Met → Val
	216, 321	4	5808	U*U*U	U*G*U	Phe → Cys
	21, 38, 661	396	4120	UA*U*	UA*C*	Tyr → Tyr
	38, 60, 661	21	4062	A*AG* AAU	A*CA* *U*AU	Lys → Ser, Asn → Tyr
	38, 60, 661	21	4063	A*AG* AAU	A*CA* *U*AU	Lys → Ser, Asn → Tyr
	38, 60, 661	21	4064	A*AG* AAU	A*CA* *U*AU	Lys → Ser, Asn → Tyr
	21, 661	396	4054	GG*G*	GG*A*	Gly → Gly
	679	23	2977	AC*C*	AC*U*	Thr → Thr
	24, 679	661	2691	A*G*G	A*AG*	Arg → Lys
	39, 661, 679	24	2573	*C*GU	*U*GU	Arg → Cys
	455, 649, 679	HR 3-9	1120	GA*A*	GA*G*	Glu → Glu
	455, 649, 679	HR 3-9	1109	*A*CU	*G*CU	Thr → Ala
	218, 455	HR 3-9	614	*A*UU	*C*UU	Ile → Leu

[a] Differences in the sequences between the heat resistant strain (HR) and the Mudd Summers strain of VSV were found in positions 1120, 1109, and 614. All other mutations were detected between different cDNA copies of the various regions of the same strain. The sequence of 011-2 covering position 6087 was previously determined from a partial cDNA clone of DI 011 (56), a DI particle RNA of the Mudd Summers parental virus.

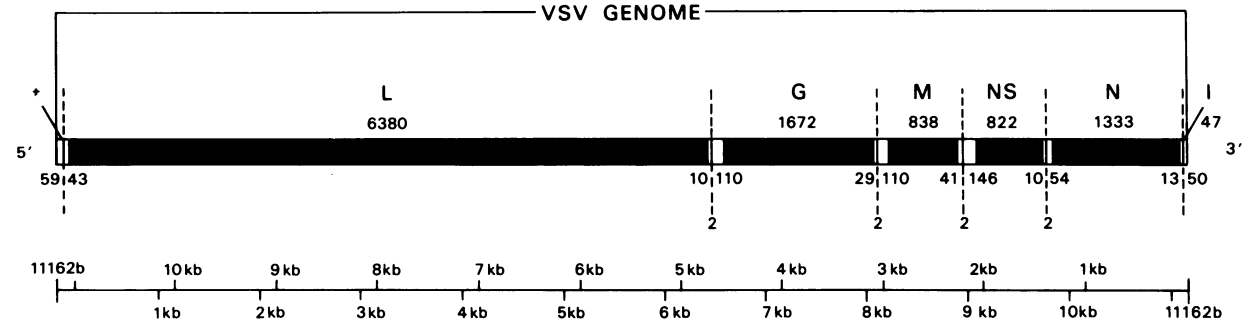

FIG. 5. VSV genome. The nucleotide sequence analysis of the *L* gene, together with the previously reported sequences of the leader region (l), the *N*, *NS*, *M*, and *G*, genes, the four two-nucleotide intercistronic regions, and the trailer region (t), completes the sequence of the VSV genome. The coding regions of each gene are marked by closed boxes. The sizes of the messages without poly(A) tails and the terminal untranslated regions of each message are indicated above and below, respectively. The genome of VSV is 11,162 nucleotides long, of which the *L* gene alone comprises 57.4%. Of the nucleotides in the genome, 99.4% are transcribed, and 93.9% are translated.

ribosome binding site which covers the first AUG of the message (43) and the polyadenylation site of *L* (47) have been identified earlier, it is likely that the detected single open reading frame of *L* mRNA is actually used for translation (Fig. 3). There is no evidence for splicing of any of the VSV messages or of messages from any other nonsegmented negative-strand RNA viruses. Therefore, we conclude that the open reading frame codes for a single protein 2,109 amino acids long with a molecular weight of 241,012. The molecular weight of the L protein was estimated earlier, varying from 160,000 to 230,000 (24, 52), but exact size markers were not available. At this point, we cannot rule out the possibility of posttranslational processing of the protein, which may lead to a substantial reduction of its size. Resolution of this possibility will have to await isolation of monospecific antibodies directed against the amino and carboxy termini of the protein. There is no evidence of other posttranslational modifications of the protein such as phosphorylation, glycosylation, or acetylation.

The size of the L protein, which is 10 times larger than the NS protein, makes it a perfect candidate for a multifunctional protein. The composition and the stoichiometry of the polymerase complex are completely unknown. Intragenic complementation among *L* gene mutants (5, 14, 15) raises the possibility that the complex may contain more than one L subunit. The presence and the role of the NS protein within the complex is totally obscure, especially since at least 13 different forms of NS can be separated which differ in their degree of phosphorylation (22). It is our hope that the two new approaches that are now possible, site-specific mutations and the use of monospecific antibodies, will enable us to study the individual functions of the polymerase complex and its interactions with the ribonucleoprotein complex.

Effects of *L* gene mutations. The 17 sequence differences between cDNA copies derived from RNA of the Mudd Summers strain (a total of 17,800 bases) suggest a frequency of mutations of approximately 10^{-3}. Between the HR and Mudd Summers strains, three sequence differences were found in 1,140 nucleotides, corresponding to a frequency of roughly 2.5×10^{-3}. As expected, the difference between strains appears greater than that among genomes of the same strain. In comparison, the error rate of the reverse transcriptase in vitro with single-stranded ϕ X DNA as a template was averaged to approximately 10^{-3} (17). With influenza virus RNA used as a more natural template, the accuracy was estimated to be slightly higher, 3×10^{-4} (12). Both values are roughly comparable to the frequency of mutations detected within the *L* gene clones, considering that incubation conditions and different preparations of the reverse transcriptase may also influence the fidelity of reverse transcription. The exact error rate of reverse transcription with an RNA template is unknown and may be overestimated since RNA templates themselves may already be heterogeneous. We cannot rule out at this time the possibility that some or even most of the sequence heterogeneities may be due to errors by the reverse transcriptase. However, clones 21 and 396, which both showed a single-base deletion in the exact same position, clearly rule out a random mistake by the reverse transcriptase. A single-base deletion was previously detected during the sequence analysis of the *HA* gene of influenza virus (6). Here, we observed two cDNA copies with the identical deletion, strongly suggesting that a subpopulation of VSV genomic RNAs may contain this deletion.

Although we do not know what influence the amino acid substitutions listed in Table 2 would have on the performance of the polymerase and its multiple functions, we can clearly evaluate the effects of single-base insertions or deletions which would lead to the translation of an incomplete L protein. Even in the case of clone 455, in which the deletion and the insertion 187 bases downstream of the message could potentially result in the restoration of the reading frame, a complete L protein would not be synthesized since translational stop codons were present in the other reading frames between the sites of deletion and insertion. It is not known whether portions of the amino terminus of the L protein representing only 27% of the protein are capable of transcription and replication, but we consider this a very remote possibility. Therefore, the frame shift mutations within the *L* gene would create a new group of defective particles which can only be propagated in the presence of parental helper virus. We expect that they would interfere with the replication of the parental virus like the transcribing DI particle LT. In addition, the absence of selective pressures on these DI particle genomes would allow accumulation of additional mutations within these genomes. It is conceivable, for example, that the RNA genomes from which clones 21 and 396 were reverse transcribed share a progenitor genome with a single-base deletion and that the additional mutations in positions 4120 and 4054 were introduced at a later stage.

VSV polymerase gene as a mutator gene. Effects of the high

mutation rates of RNA viruses, in particular of influenza virus, have been painfully experienced throughout centuries. The high mutability, perhaps owing to the absence of proofreading mechanisms, exerts itself most dramatically in numerous epidemic outbreaks caused by the antigenic drift of new variants of the same virus genus and the lack of immunity in, for example, the human population. The genetics and variability of VSV have been extensively studied in many laboratories. In this communication, we presented direct sequencing evidence of the variability of VSV genomes, particularly the variability of the *L* gene. However, it is possible that some or most of the sequence heterogeneities may have been introduced by errors of the reverse transcriptase reactions. The high number of differences found within the polymerase gene (Fig. 4, Table 2) was expected between two different strains (HR and Mudd Summers). The surprisingly large variability between cDNA copies of RNA of the same strain, which after plaque purification was passaged two times, suggests that every single RNA genome may be different. Since the polymerase gene represents 57% of the VSV genome, a large number of these mutations should fall within this gene. In fact, the viral polymerase itself may be a mutator protein, and the mutations found within the *L* gene may affect the fidelity of the polymerase and thereby the rate at which mutations are introduced. Since the polymerase complex consists of not only the L but also the NS protein, which together specifically interact with the ribonucleoprotein template, mutational changes in any of these three components may lead to an increased error rate. Mistakes made by the polymerase during replication consequently contain an element of positive feedback: mutations in the polymerase gene and also the *NS* or *N* genes can be expected to increase the probability of subsequent mistakes, a hypothesis originally formulated by Orgel (36, 37) to explain senescense based on an error catastrophe during protein translation. This positive feedback may, in part, support the maintenance of persistent viral infections. In contrast to acute VSV infections (7), *ts* mutants with an RNA minus phenotype are rapidly selected during persistent infections (38). In addition, the genomes rapidly and continuously evolve, as measured by T1 oligonucleotide mapping (20). It therefore appears that besides its multiple functions during transcription and replication, the VSV polymerase gene plays an important role as a mutator gene in the rapid evolution of the virus and that the erosion of its fidelity may contribute to maintaining the persistent state of an infection.

ACKNOWLEDGMENTS

We are indebted to Robert A. Lazzarini for his constructive comments on this research and to Lynn Hudson and Judy Sprague for allowing us access to their VSV cDNA libraries. We also thank Charlene French and David Powell for their continuous support in computer programming and analysis of the *L* gene sequences. We thank Jacob V. Maizel and John Owens for comparing the *L* gene sequences with other sequences.

LITERATURE CITED

1. **Abraham, G., and A. K. Banerjee.** 1976. Sequential transcription of the genes of vesicular stomatitis virus. Proc. Natl. Acad. Sci. U.S.A. **73**:1504–1508.
2. **Ball, L. A., and G. W. Wertz.** 1981. VSV RNA synthesis: how can you be positive? Cell **26**:143–144.
3. **Ball, L. A., and C. N. White.** 1976. Order of transcription of genes of vesicular stomatitis virus. Proc. Natl. Acad. Sci. U.S.A. **73**:442–446.
4. **Banerjee, A. K.** 1980. The *in vitro* mRNA transcription process. p. 35–50. *In* D. H. L. Bishop (ed.), Rhabdoviruses, vol. 2. CRC Press, Boca Raton, Fla.
5. **Belle Isle, H. D., and S. U. Emerson.** 1982. Use of a hybrid infectivity assay to analyze primary transcription of temperature-sensitive mutants of the New Jersey serotype of vesicular stomatitis virus. J. Virol. **43**:37–40.
6. **Both, G. W., and M. J. Sleigh.** 1980. Complete nucleotide sequence of the haemagglutinin gene from a human influenza virus of the Hong Kong subtype. Nucleic Acids Res. **8**:2561–2575.
7. **Clewley, J. P., D. H. L. Bishop, C.-Y. Kang, J. Coffin, W. M. Schnitzlein, M. E. Reichmann, and R. E. Shope.** 1977. Oligonucleotide fingerprints of RNA species obtained from rhabdoviruses belonging to the vesicular stomatitis subgroup. J. Virol. **23**:152–166.
8. **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.
9. **Deng, G., and R. Wu.** 1981. An improved procedure for utilizing terminal transferase to add homopolymers to the 3′ termini of DNA. Nucleic Acids Res. **9**:4173–4188.
10. **Donis-Keller, H., A. M. Maxam, and W. Gilbert.** 1977. Mapping adenines, guanines and pyrimidines in RNA. Nucleic Acids Res. **4**:2527–2538.
11. **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.
12. **Fields, S., and G. Winter.** 1981. Nucleotide sequence heterogeneity and sequence rearrangements in influenza virus cDNA. Gene **15**:207–214.
13. **Fields, S., and G. Winter.** 1982. Nucleotide sequences of influenza virus segments 1 and 3 reveal mosaic structure of a small viral RNA segment. Cell **28**:303–313.
14. **Flamand, A.** 1970. Étude génétique du virus de la stomatite vésiculaire: classement de mutants thermosensibles spontanés en groupes de complementation. J. Gen. Virol. **8**:187–195.
15. **Gadkari, D. A., and C. R. Pringle.** 1980. Temperature-sensitive mutants of Chandipura virus. I. Inter- and intra-group complementation. J. Virol. **33**:100–106.
16. **Gallione, C. J., J. R. Greene, L. E. Iverson, and J. K. Rose.** 1981. Nucleotide sequences of the mRNA's encoding the vesicular stomatitis virus N and NS proteins. J. Virol. **39**:529–535.
17. **Gopinathan, K. P., L. A. Weymouth, T. A. Kunkel, and L. A. Loeb.** 1979. Mutagenesis *in vitro* by DNA polymerase from an RNA tumour virus. Nature (London) **278**:857–859.
18. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. **72**:3961–3965.
19. **Herman, R. C., M. Schubert, J. D. Keene, and R. A. Lazzarini.** 1980. Polycistronic vesicular stomatitis virus RNA transcripts. Proc. Natl. Acad. Sci. U.S.A. **77**:4662–4665.
20. **Holland, J., E. Grabau, C. Jones, and B. Semler.** 1979. Evolution of multiple genome mutations during long-term persistent infection by vesicular stomatitis virus. Cell **16**:495–504.
21. **Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. Vande Pol.** 1982. Rapid evolution of RNA genomes. Science **215**:1577–1585.
22. **Hsu, C.-S., E. M. Morgan, and D. W. Kingsbury.** 1982. Site-specific phosphorylation regulates the transcriptive activity of vesicular stomatitis virus NS protein. J. Virol. **43**:104–112.
23. **Huang, A. S., and E. K. Manders.** 1972. Ribonucleic acid synthesis of vesicular stomatitis virus. IV. Transcription by standard virus in the presence of defective interfering particles. J. Virol. **9**:909–916.
24. **Kang, C. Y., and L. Prevec.** 1969. Proteins of vesicular stomatitis virus. I. Polyacrylamide gel analysis of viral antigens. J. Virol. **3**:404–413.
25. **Keene, J. D., I. M. Chien, and R. A. Lazzarini.** 1981. Vesicular stomatitis virus defective particle contains a muted internal leader RNA gene. Proc. Natl. Acad. Sci. U.S.A. **18**:2090–2094.
26. **Keene, J. D., M. Schubert, and R. A. Lazzarini.** 1979. Terminal sequences of vesicular stomatitis virus RNA are both complementary and conserved. J. Virol. **32**:167–174.

27. **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.
28. **Keene, J. D., M. Schubert, R. A. Lazzarini, and M. Rosenberg.** 1978. Nucleotide sequence homology at the 3′ termini of RNA from vesicular stomatitis virus and its defective interfering particles. Proc. Natl. Acad. Sci. U.S.A. **75**:3225–3229.
29. **Land, H., M. Grez, H. Hauser, W. Lindenmaier, and G. Schuetz.** 1981. 5′-terminal sequences of eucaryotic mRNA can be cloned with high efficiency. Nucleic Acids Res. **9**:2251–2266.
30. **Lazzarini, R. A., J. D. Keene, and M. Schubert.** 1981. The origins of defective interfering particles of the negative-strand RNA viruses. Cell **26**:145–154.
31. **Lazzarini, R. A., G. H. Weber, L. D. Johnson, and G. M. Stamminger.** 1975. Covalently linked message and anti-message (genomic) RNA from a defective vesicular stomatitis virus particle. J. Mol. Biol. **97**:289–307.
32. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
33. **McGeoch, D. J.** 1979. Structure of the gene N: gene NS intercistronic junction in the genome of vesicular stomatitis virus. Cell **17**:673–681.
34. **McGeoch, D. J., and A. Dolan.** 1979. Sequence of 200 nucleotides at the 3′ terminus of the RNA genome of vesicular stomatitis virus. Nucleic Acids Res. **6**:3199–3211.
34a. **Meier, E., G. G. Harmison, J. D. Keene, and M. Schubert.** 1984. Sites of copy choice replication involved in generation of vesicular stomatitis virus defective-interfering particle RNAs. J. Virol. **51**:515–521.
35. **Murphy, M. F., and R. A. Lazzarini.** 1974. Synthesis of viral mRNA and polyadenylate by a ribonucleoprotein complex from an extract of VSV-infected cells. Cell **3**:77–84.
36. **Orgel, L. E.** 1963. The maintenance of the accuracy of protein synthesis and its relevance to aging. Proc. Natl. Acad. Sci. U.S.A. **49**:517–521.
37. **Orgel, L. E.** 1970. The maintenance of the accuracy of protein synthesis and its relevance to aging: a correction. Proc. Natl. Acad. Sci. U.S.A. **67**:1476.
38. **Preble, O. T., and J. S. Youngner.** 1972. Temperature-sensitive mutants isolated from L cells persistently infected with Newcastle disease virus. J. Virol. **9**:200–206.
39. **Pringle, C. R., I. B. Duncan, and M. Stevenson.** 1971. Isolation and characterization of temperature-sensitive mutants of vesicular stomatitis virus, New Jersey serotype. J. Virol. **8**:836–841.
40. **Queen, C., and L. J. Korn.** 1980. Computer analysis of nucleic acids and proteins. Methods Enzymol. **65**:595–609.
41. **Repik, P., and D. H. L. Bishop.** 1973. Determination of the molecular weight of animal RNA viral genomes by nuclease digestions. I. Vesicular stomatitis virus and its defective T particle. J. Virol. **12**:969–983.
42. **Richardson, C. C.** 1965. Phosphorylation of nucleic acid by an enzyme from T4 bacteriophage-infected Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. **54**:158–165.
43. **Rose, J. K.** 1978. Complete sequences of the ribosome recognition sites in vesicular stomatitis virus mRNAs: recognition by the 40S and 80S complexes. Cell **14**:345–353.
44. **Rose, J. K.** 1980. Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. Cell **19**:415–421.
45. **Rose, J. K., and C. J. Gallione.** 1981. Nucleotide sequence 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.
46. **Rowlands, D. J.** 1979. Sequences of vesicular stomatitis virus RNA in the region coding for leader RNA, N protein mRNA, and their junction. Proc. Natl. Acad. Sci. U.S.A. **76**:4793–4797.
47. **Schubert, M., J. D. Keene, R. C. Herman, and R. A. Lazzarini.** 1980. Site on the vesicular stomatitis virus genome specifying polyadenylation and the end of the L gene mRNA. J. Virol. **34**:550–559.
48. **Schubert, M., and R. A. Lazzarini.** 1981. In vivo transcription of the 5′-terminal extracistronic region of vesicular stomatitis virus RNA. J. Virol. **38**:256–262.
49. **Sprague, J., J. H. Condra, H. Arnheiter, and R. A. Lazzarini.** 1983. Expression of a recombinant DNA gene coding for the vesicular stomatitis virus nucleocapsid protein. J. Virol. **45**:773–781.
50. **Tu, C. P. D., and S. N. Cohen.** 1980. 3′-end labeling of DNA with (α-^{32}P)cordycepin-5-triphosphate. Gene **10**:177–183.
51. **Ulmanen, J., B. A. Broni, and R. M. Krug.** 1981. Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m^7GpppNm) on RNAs and in initiating viral RNA transcription. Proc. Natl. Acad. Sci. U.S.A. **78**:7355–7359.
52. **Wagner, R. R.** 1975. Reproduction of Rhabdoviruses, p. 1–93. *In* H. Fraenkel-Conrat and R. R. Wagner (ed.), Comprehensive virology, vol. 4. Plenum Publishing Corp., New York.
53. **Wahli, W., G. U. Ryffel, T. Wyler, R. B. Jaggi, R. Weber, and I. B. David.** 1978. Cloning and characterization of synthetic sequences from the xenopus laevis vitellogenin structural gene. Dev. Biol. **67**:371–383.
54. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. U.S.A. **80**:726–730.
55. **Winter, G., and S. Fields.** 1982. Nucleotide sequence of human influenza A/PR/8/34 segment 2. Nucleic Acids Res. **10**:2135–2143.
56. **Yang, F., and R. A. Lazzarini.** 1983. Analysis of the recombination event generating a vesicular stomatitis virus deletion defective interfering particle. J. Virol. **45**:766–772.
57. **Youngner, J. S., and D. O. Quagliana.** 1976. Temperature-sensitive mutants of vesicular stomatitis virus are conditionally defective particles that interfere with and are rescued by wild-type virus. J. Virol. **19**:102–107.