Sequence signal involved in the generation of an internally deleted defective interfering RNA from vesicular stomatitis virus

Barun K. De and Jacques Perrault

Department of Microbiology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA

Received 17 June 1982; Revised and Accepted 13 September 1982

ABSTRACT

We have determined the nucleotide sequence at the 3' end of the vesicular stomatitis virus (VSV) polymerase or L gene and compared it to that obtained from a defective interfering particle (DI) RNA generated by this virus. The latter (DI 0.50) contains a large internal deletion within this gene. The deletion begins exactly 253 residues from the transcription start of the gene and extends to ~300 bases from the 5' end of the standard RNA genome. The flanking sequences bear no homology to the eukaryotic consensus splice sequence. The sequence immediately preceding the deletion is complementary to the ribosome binding site of the L gene transcript and also resembles genome transcription termination sites. In addition, we present the results of nuclease protection experiments which show that this DI RNA retains an exact copy of the 3' end half of the standard genome (leader, N, NS, M and G genes), although its own 3' end is non-genomic. The implications of these findings regarding mechanisms of DI generation are discussed.

INTRODUCTION

Defective interfering particles (DI) of RNA viruses represent an important class of virus regulatory elements. They interfere strongly with their parent helper virus replication and play a major role in the establishment and maintenance of persistent viral infections (for review see 1 and 2).

The deleted genomes characteristic of DI have been studied most extensively in the vesicular stomatitis virus system. Mechanisms involved in the generation of the various genomic rearrangements observed in negative-strand virus DI have recently been reviewed (1,3). "Copy-back" sequences (short stretches complementary to the 5' end of the genome) are commonly found at the 3' ends of VSV DI RNA molecules which otherwise represent various size segments extending from the 5' end of the genome. These "copy-back" sequences are thought to be crucial for the interference property of most VSV DI. At least two DI derived from the VSV HR strain, on the other hand, contain internal genome deletions (4,5,6). One of these is a simple internal deletion (DI 0.46, also called DI-LT) while the other (DI 0.50

or DI-LT₂) contains both an internal deletion and a 3' end copy-back sequence. Only the former is transcriptionally active and displays biological properties not shared with the other major class of VSV DI (4,7). We and others have presented evidence that the internal deletion in VSV DI 0.50 is situated within the large polymerase or L gene. We present here the results of nucleotide sequencing studies spanning the DI 0.50 deletion as well as nuclease protection experiments indicative of the overall sequence organization of this DI RNA molecule. These data support the hypothesis that specific template sequences may be involved in the generation of internal deletions in VSV DI.

MATERIALS AND METHODS

Preparation of template RNAs

Growth of BHK₂₁ host cells, standard virus DI 0.50 and DI 0.46, ³H-uridine labeling, virus and DI purification, and RNA extractions were described previously (4,8,9).

Preparation of single-stranded cDNA primer

The plasmid pG65 (a gift from Dr. John Rose, Salk Institute) was constructed from pBR322 by inserting a sequence corresponding to the 3' terminus of the VSV G mRNA at the Pst I site (10). Growth and purification of the plasmid followed published procedures (11). Isolation of the 5' end-labeled primer was carried out as described by Rose (12) except for strand separation which we carried out by denaturing the sample in 80% dimethyl sulfoxide for 2 min at 85°C, quick cooling, and immediate electrophoresis in a non-denaturing acrylamide gel. Recovery of the single-stranded 49 base long primer (lower band) by gel crushing and shaking at 37°C overnight, was followed by loading onto a 1 ml DE52 column, washing with several volumes of 0.1 M NaCl, 10 mM Tris-Cl pH 8.0 buffer, and elution with the same buffer containing 1 M NaCl.

Primer extension and sequencing

The procedure of Rose (12) was followed except for the following modifications. Templates (10–15 µg of standard genome or DI 0.50 RNA) and about equimolar amounts of primer (40–50 ng, ~2 x 10⁶ cpm) were first suspended in 50 µl of 50 mM Tris-Cl, pH 8.4, 30 mM β-mercaptoethanol, 110 mM KCl and heated to 85°C for 1 min, and then incubated at 42°C for 2 h. Reaction volumes were adjusted to 100 µl with final concentrations of 10 mM MgCl₂, 4 mM sodium pyrophosphate, 0.5 mM of the four dNTPs, 300–400 units/ml of reverse transcriptase (a gift from Dr. J.W. Beard), 50 mM Tris-Cl, pH 8.4,

30 mM β -mercaptoethanol, and 110 mM KCl. After 2 h at 42°C the reaction was terminated with EDTA (25 mM final concentration), phenol extracted and ethanol precipitated. The RNA strand was hydrolyzed with 0.2 M NaOH at 65°C for 30 min, the mixture neutralized with HCl, and the extended products purified over a G-50 Sephadex column before final ethanol precipitation. Conversion of the 5' end labeled cDNA primer to long products (>50% larger than 300 bases) was first verified by denaturing gel analysis before carrying out chemical sequencing.

Cleavage reactions for sequencing were carried out as described by Maxam and Gilbert (13) with the modification of Smith and Calvo (14). Sequencing gels of 6.5% or 5% total acrylamide concentration (30 x 80 x 0.05 cm) were prepared and run as described by Sanger and Coulson (15).

Nuclease protection experiments

Self-hybridization of 0.50 RNA or hybridization of radiolabeled 0.46 RNA to 0.50 RNA was carried in 0.3 M or 0.6 M NaCl, 10 mM Tris-Cl, pH 7.8, 1 mM EDTA, at 70°C, and at the concentrations and for the times indicated. Samples were first denatured at 100°C in the above buffer without NaCl. Nuclease digestion was then carried out in the same hybridization buffer at 37°C for 30 min, with 5 μ g/ml and 2.5 units/ml of ribonuclease A and T1 respectively, or four times the concentration of nucleases in the case of 3' end-labeled 0.46 RNA annealed to fragmented, double-stranded 0.50 RNA. The latter was prepared by self-annealing of a 60 μ g sample in 200 μ l for 15 h, digestion with the 4x concentration of nucleases, followed by alkaline fragmentation in 0.1 M NaCO₃ for 30 min at 50°C. For recovery of the double-stranded fraction of 0.50 RNA with minimal fragmentation, hybridization was carried out at 600 μ g/ml for 15 min followed by digestion with the lower amounts of nuclease. Hybridized digested samples were always recovered by ethanol precipitation following proteinase K digestion and phenol-chloroform extraction (4). 3' end-labeling of 0.46 RNA and 5' end-labeling of 0.50 RNA and analysis of glyoxal-denatured RNAs in agarose gels or undenatured duplexes in 20% acrylamide gels, were carried out as described previously (4).

RESULTS

Sequencing strategy

To examine those sequences involved in the generation of internal deletions in VSV DI we determined the nucleotide sequence of the relevant template regions. Previous studies suggested that the DI 0.50 RNA deletion is located close to the 5' end of the G gene (4,5). Except for minor modifications, our

Fig. 1. Sequencing strategy. The nucleotide sequence in the region spanning the G-L gene junction and extending into the VSV polymerase gene was determined for both the standard genome RNA and its DI 0.50. The 5' end-labeled cDNA primer was annealed to the templates and extended with reverse transcriptase. The sequence of extended products was determined as described in Materials and Methods. The shaded box within the L gene corresponds roughly to the sequence deleted in DI 0.50.

methodology followed the procedure of Rose (12) who determined the G-L gene junction sequence by primer extension of cDNA annealed to genome RNA (Fig. 1). The 5' end-labeled cDNA primer (49 nucleotides complementary to bases 33 to 81 from the end of the 1665 nucleotide long G gene) was derived from the pG65 plasmid (10). The cDNA nucleotide sequence of the extended products was determined by the method of Maxam and Gilbert (13). Several independent sequence determinations were carried out with different template RNA preparations to ensure accuracy.

Sequences from the L gene of the standard VSV genome and DI 0.50 RNA

The sequence of the first 45 nucleotides of the L gene previously determined for the San Juan VSV Indiana strain (12,16) is in complete agreement with our results with the standard VSV HR Indiana strain (Fig. 2). The remainder of the sequence which could be unambiguously read until residue 306 retains the opened reading frame starting with the AUG codon at position 11. Start codons in phase with the second and third reading frames are all followed by stop codons within a short stretch. The predicted L protein NH₂ terminal amino acid sequence is also shown.

The nucleotide sequence of the DI 0.50 template RNA is completely homologous to that of the standard genome template up to and including residue 252. It then diverges for the remainder of the sequence determined (Fig. 2). It should be noted that our ability to read a clear-cut sequence for the DI 0.50 template attests to the relative homogeneity of this RNA preparation at least for the region in question. However, we could not resolve the identity of certain nucleotides at a few specific sites (see Fig. 2) because bands were present in all lanes of the sequencing gel. Fig. 3 illustrates this problem for the ambiguities adjacent to the site of deletion (see Discussion).

Overall genome sequence rearrangements in DI 0.50 RNA

We have previously demonstrated the presence of the four 3' end proximal genes of the VSV genome (N, NS, M and G) within DI 0.50 as well as a

Fig. 2. Nucleotide sequence at the beginning of the L gene in the standard VSV genome (upper) and its DI 0.50 (lower). This cDNA sequence corresponds to mRNA polarity and the corresponding amino acid sequence is also shown. The site of internal deletion in DI 0.50 RNA begins at position 253. The hexamer sequence homology between the site of deletion and the translation start as well as the repeat sequence flanking residue 49 are underlined (see Discussion). Y denotes a pyrimidine residue while X stands for undetermined.

non-genomic copy-back sequence of ~ 65 bases at the 3' end of the molecule (4). We present here the results of nuclease protection experiments which show that DI 0.50 RNA contains within itself an exact contiguous copy (including the 3' terminal residue) of the 3' half of the standard genome. As a probe for the 3' terminal half of the standard genome, we have employed DI

Fig. 3. Radioautographs of sequencing gel derived from the standard VSV genome versus the DI 0.50 template around the site of deletion (see Fig. 2).

0.46 RNA which retains this segment intact and unrearranged up to and including the G gene coding sequences (4,17,18). We have also shown previously that, in contrast to DI 0.46, DI 0.50 RNA preparations self-anneal to ~30–40% nuclease resistance (4). If this does reflect the presence of ~15–20% plus strand copies of DI 0.50 RNA as often observed in other VSV DI preparations (8) then these plus strands can be annealed to labeled 0.46 RNA to determine the extent of uninterrupted complementarity by nuclease protection. Fig. 4 shows the results of such an experiment. The nuclease-treated, self-annealed fraction of DI 0.50 RNA is identical in size to untreated 0.50 RNA as predicted (Fig. 4A). When DI 0.50 RNA is annealed to uniformly-labeled 0.46 RNA, the nuclease protected fragments include molecules ~95% as large as 0.46 RNA itself (Fig. 4B). Control annealing reactions with labeled 0.46 RNA only (~2% resistance) yield no distinct bands (not shown). These results show clearly that DI 0.50 RNA molecules contain ~95% of the DI 0.46 RNA sequences as a contiguous segment.

To establish whether DI 0.50 RNA contains the exact 3' end terminal sequence of the standard genome RNA, we carried out nuclease protection experiments using 3' end-labeled 0.46 RNA annealed to DI 0.50 RNA. In this case, we sought to increase the extent of annealing between DI 0.50 plus strands and the labeled probe by reducing the amount of competing DI 0.50 minus strands. This was accomplished by first isolating the nuclease-treated, self-annealed fraction of DI 0.50 RNA. In addition, we fragmented this RNA with alkali (see Materials and Methods) to a size range (~100–200 bases) roughly equivalent to that of the 3' end-labeled 0.46 RNA which was

Fig. 4. Nuclease protection of ^3H -uridine-labeled DI 0.50 and 0.46 RNAs. (A) lane a, untreated DI 0.50 RNA control (40 k cpm/ μg); lane b, DI 0.50 RNA (100 $\mu\text{g}/\text{ml}$) self-annealed for 20 min and nuclease digested (~30% resistant). (B) lane a, ^3H -uridine HeLa cell ribosomal RNA markers; lane b, untreated DI 0.46 RNA control (~200 k cpm/ μg); lane c, 0.46 RNA (6 $\mu\text{g}/\text{ml}$) hybridized to unlabeled, unfractionated 0.50 RNA (200 $\mu\text{g}/\text{ml}$) for 15 min, and nuclease treated (~15% resistant). Other hybridization conditions, RNase digestions, recovery of protected fragments, glyoxal denaturation, and analysis on agarose gels are described in Materials and Methods. Arrows in (A) indicate positions of marker ribosomal RNAs.

extensively nicked during the end-labeling reaction. This latter step was carried out to avoid annealing competition based on the size of the duplex molecules. Under these conditions at least 30 to 40% of the end label in 0.46 RNA (background subtracted) can be protected from nuclease digestion by DI 0.50 plus strands. The expected size heterogeneity of the protected fragments is shown in Fig. 5A. No protected labeled bands were seen in control reactions without DI 0.50 RNA (not shown). These results strongly support our previous hypothesis that DI 0.50 RNA contains the 3' end terminal sequence of the standard genome at a short distance from its own 3' end (4).

Lastly, we wished to determine the sequence arrangement at the 5' end of DI 0.50 RNA. Our previous studies established that at least the first ~55 bases of this template (the length of the hybridization probe used) are identical to the standard genome (4). In the following experiment we extended these results by using 5' end-labeled standard genome RNA as a probe for nuclease protection experiments with DI 0.50 RNA. In this instance, the self-annealed fraction of DI 0.50 RNA was isolated by gentle nuclease treatment to preserve plus strands as intact as possible. Fig. 5B shows the size of the two 5' end-labeled genome segments protected from nuclease digestion (the sample showed ~25% resistance above background). The small

Fig. 5. Nuclease protection of end-labeled DI 0.46 and standard VSV genome RNAs. (A) Protected 3' end-32p-labeled duplex RNAs were analyzed on a 20% acrylamide gel under non-denaturing conditions. Lane a, 48 bp RNA marker obtained by hybridization of 3' end-labeled 0.46 RNA to VSV transcription products (see 4); lane b, 3' end labeled 0.46 RNA (5 k cpm/ μ g, 5 μ g/ml) was annealed to the fragmented, self-annealed fraction of DI 0.50 RNA (100 μ g/ml) for 140 min and nuclease digested (see text). (B) Protected fragments of 5' end-32p-labeled standard VSV were analyzed on agarose gels following nuclease digestion and glyoxal denaturation (see Fig. 1). 5' end-labeled standard RNA (~40 μ g/ml, ~5 k cpm/ μ g) was hybridized to the unlabeled self-annealed fraction of 0.50 RNA (~90 μ g/ml) for 20 min and nuclease digested (see text). Arrows indicate positions of 32p-labeled HeLa ribosomal RNA markers.

Fig. 6. Genome sequence rearrangements in VSV DI 0.50. The genome is drawn to scale with the regions flanking the deletion within the L gene and the rearranged 3' end magnified to show details.

Fig. 7. Sequence homologies between the DI 0.50 internal deletion site and putative genome transcription termination signals. Arrows indicate the sites of termination at the VSV gene junctions and within the N gene as well as the deletion site in DI 0.50 RNA. The intercistronic dinucleotide common to all VSV gene junctions (except NS-M which is CA) is underlined. The observed homologies are boxed in (see text).

protected species corresponds in size to the ~65 base long terminal complementary sequences of DI 0.50, which are present in the nuclease-treated self-annealed fraction of DI 0.50 (see Fig. 4A, lane b). The larger protected segment, ~300 bases long, must therefore correspond to the contiguous 5' end genome segment present in DI 0.50 minus strands. Control annealing reactions without DI 0.50 contained only the labeled material which failed to enter the gel (not shown). These results strongly support our previous suggestion, based on MW comparisons between DI 0.46 and DI 0.50 RNAs, that the latter RNA contains a single internal deletion within the L gene (see Fig. 6).

DISCUSSION

The VSV genome sequence rearrangements in DI 0.50 RNA are illustrated in Fig. 6. As suggested previously (1) this DI RNA most likely corresponds to the DI-LT₂ characterized by Lazzarini and co-workers (5,7). Both DI preparations are derived from the same HR-LT DI isolate (19), are comparable in size, and share properties so far unique in VSV DI, namely, the presence of both an internal deletion and a copy-back sequence. Sequencing results have clearly established that the 3' end of DI-LT₂ is a 70 base long copy-back sequence (complementary to the first 70 nucleotides at the 5' end), followed by one non-coded A residue, and then followed by the exact 3' end terminal sequence of the standard genome (7). The nuclease protection results presented in this communication establish that this 3' end sequence arrangement is also characteristic of DI 0.50, as suggested earlier (4). In

addition, our experiments clearly show that no other 3' end genome sequence rearrangements occur in DI 0.50 within limits of detection of the nuclease protection assay.

Our sequence determination of the beginning of the L gene from both the standard and DI 0.50 RNA templates is unambiguous except at a few selected sites (see Figs. 2 and 3). Possible explanations for these ambiguities include technical difficulties such as band compression, or preferential stops during primer extension, or mutational heterogeneity (20). We cannot at present exclude any of these possibilities.

The most significant aspect of our results deals with the question of possible sequence specificity at the sites of VSV internal deletions. In support of generation models (1,3) based on viral replicase "errors", we find no obvious homology between the site of the DI 0.50 deletion and the eukaryotic consensus splice sequence. The same conclusion has been reached by others regarding internal deletions in togavirus DI RNAs (21,22) and influenza DI RNAs (23,24). More importantly, a closer examination of the nucleotide sequence flanking the DI 0.50 deletion reveals homologies to sequences at or near VSV transcription termination sites (Fig. 7). The sequence 3'-AUUGU-5', which includes one of the two conserved and non-transcribed intercistronic nucleotides as well as the common transcription start sequence (12), is found only five residues upstream from the deletion start. Furthermore, the sequence 3'-UACAG-5' spanning the deletion site is also found at the site of premature termination in the N gene where the majority of initiated transcripts represent uncapped oligonucleotides 12 to 14 residues in length (25). The synthesis of these appears to be dependent on M protein binding to the ribonucleoprotein template (26). No evidence has yet been obtained for an analogous premature termination site for L gene transcripts, but the last six nucleotides preceding the DI 0.50 deletion are also homologous to residues 8 to 13 from the start of the gene (Fig. 2).

It is not clear as yet which nucleotide sequences are essential for transcription termination and/or M protein binding but the close proximity of termination and start sites may reflect an overlap of sequence signals. In any case, it seems plausible that VSV genome sequences bearing homology to intercistronic borders could trigger the requisite interruption of viral RNA synthesis for the generation of DI molecules. Note that VSV DI generation occurs at a frequency of about 10^{-7} to 10^{-8} per infectious particle replication (1). Such low frequency events may reflect in part a low probability of termination when the VSV polymerase encounters a template

sequence bearing only partial homology to normal transcription termination signals. Furthermore, such imperfect signals may be expected to occur at a multiplicity of sites throughout the VSV genome or its complement (for e.g., a given pentamer sequence should occur about ten times in a single-stranded genome of 11 kb on a random basis) which is consistent with the apparent diversity of deletion sites observed in different VSV DI (1). A recent study also proposes that homology to transcription termination sites may be responsible for the generation of complex DI genome rearrangements during plus strand synthesis of influenza virus (23).

Lastly, we mention two important features of the VSV L gene sequence determined. First, the perfect eight nucleotide repeat sequence (beginning at residue 41 and interrupted by one residue at position 49, see Fig. 2) may not be fortuitous (chance probability of 0.17 for a genome of 11 kb) and could perhaps represent an important regulatory signal. Second, we also note an unusually acidic domain at the putative amino terminal end of the L polymerase protein (9 acidic and no basic residues within the first 18 amino acids). An analogous acidic region is also found within the body of the VSV NS protein (27).

ACKNOWLEDGEMENTS

We wish to thank Dr. Jack Rose for providing the pG65 plasmid used in this work, Dr. Marcella McClure for many helpful discussions and suggestions, Drs. Sondra Schlesinger, Stuart Nichol and Donald Rao for critical reading of the manuscript, and Joanne Shuttleworth for expert technical assistance. This work was supported by NIH grant AI-14365 and a Research Career Development Award to J.P. and a Berg Fellowship to B.D.

REFERENCES

1. Perrault, J. (1981) in *Initiation Signals in Viral Gene Expression*, Shatkin, A.J. Ed., Curr. Topics Microbiol. Immunol. 93, 151-207.
2. Holland, J.J., Kennedy, S.I.T., Semler, B.L., Jones, C.L., Roux, L. and Grabau, E.A. (1980) in *Comprehensive Virology*, Fraenkel-Conrat, H. and Wagner, R.R. Eds., Vol. 16, pp. 137-192, Plenum Press, New York.
3. Lazzarini, R.A., Keene, J.D. and Schubert, M. (1981) *Cell* 26, 145-154.
4. Perrault, J. and Semler, B.L. (1979) *Proc. Natl. Acad. Sci. USA* 76, 6191-6195.
5. Epstein, D.A., Herman, R.C., Chien, I. and Lazzarini, R.A. (1980) *J. Virol.* 33, 818-829.
6. Clerx-Van Haaster, C.M., Clewley, J.P. and Bishop, D.H.L. (1980) *J. Virol.* 33, 807-817.
7. Keene, J.D., Chien, I.M. and Lazzarini, R.A. (1981) *Proc. Natl. Acad. Sci. USA* 78, 2090-2094.
8. Perrault, J. and Leavitt, R.W. (1977) *J. Gen. Virol.* 38, 21-34.

9. Perrault, J. and Leavitt, R.W. (1977) *J. Gen. Virol.* 38, 35-50.
10. Rose, J.K. and Iverson, L. (1979) *J. Virol.* 32, 404-411.
11. Davis, R.W., Botstein, D. and Roth, J.R. (1980) in *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
12. Rose, J.K. (1980) *Cell* 19, 415-421.
13. Maxam, A.M. and Gilbert, W. (1980) in *Methods in Enzymology*, Grossman, L. and Moldave, K. Eds., Vol. 65, part 1, pp. 499-560, Academic Press, New York.
14. Smith, D.R. and Calvo, J.M. (1980) *Nucleic Acids Res.* 8, 2255-2274.
15. Sanger, F. and Coulson, A.R. (1978) *FEBS Lett.* 87, 107-110.
16. Rose, J.K. (1978) *Cell* 14, 345-353.
17. Chow, J.M., Schnitzlein, W.M. and Reichmann, M.E. (1977) *Virology* 77, 579-588.
18. Johnson, L.D. and Lazzarini, R.A. (1978) in *Persistent Virus*. ICN-UCLA Symposia on Molecular and Cellular Biology, Stevens, J.G. and Todaro, G. J. Eds., Vol. XI, pp. 409-416, Academic Press, New York.
19. Prevec, L. and Kang, C.Y. (1970) *Nature* 228, 25-27.
20. Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. and VandePol, S. (1982) *Science* 215, 1577-1585.
21. Lethovaara, P., Soderlund, H., Keranen, S., Petterson, R.F. and Kaariainen, L. (1981) *Proc. Natl. Acad. Sci. USA* 78, 5353-5357.
22. Monroe, S.S., Ou, J-H., Rice, C.M., Schlesinger, S., Strauss, E.G. and Strauss, J.H. (1982) *J. Virol.* 41, 153-162.
23. Fields, S. and Winter, G. (1982) *Cell* 28, 303-313.
24. Nayak, D.P., Sivasubramanian, N., Davis, A.R. Cortini, R. and Sung, J. (1982) *Proc. Natl. Acad. Sci. USA* 79, 2216-2220.
25. Pinney, D.F. and Emerson, S.V. (1982) *J. Virol.* 42, 889-896.
26. Pinney, D.F. and Emerson, S.V. (1982) *J. Virol.* 42, 897-904.
27. Gallione, C.J., Greene, J.R., Iverson, L.E. and Rose, J.K. (1981) *J. Virol.* 39, 529-535.