Editing of the Sendai Virus P/C mRNA by G Insertion Occurs during mRNA Synthesis via a Virus-Encoded Activity

SILVIA VIDAL, JOSEPH CURRAN, AND DANIEL KOLAKOFSKY*

*Department of Microbiology, University of Geneva Medical School, CMU, 9 Avenue de Champel,
CH-1211 Geneva, Switzerland*

Two forms of the Sendai virus P/C mRNA have been predicted: one an exact copy of the viral genome, and the other with a single G insertion within a run of three G's. We directly cloned the mRNA or portions of it containing the insertion site and screened the resulting colonies with oligonucleotides that could distinguish the presence of three or four G's at this position. We found that 31% of the mRNAs did in fact contain the predicted insertion, whereas the viral genomes contained no heterogeneity at this position. A smaller fraction (7%) of the mRNA contained two to eight G's inserted at this position. The insertions also took place during RNA synthesis in vitro with purified virions but were not detected when the mRNA was expressed in vivo via a vaccinia virus recombinant. When the Sendai virus- and vaccinia virus-derived P/C mRNAs were coexpressed in the same cells under conditions in which each could be distinguished, those from the Sendai genome were altered as before, but those from the vaccinia virus genome remained unaltered. The activity that alters the mRNA is therefore likely to be coded for by the virus and cannot function in *trans*.

Sendai virus, a paramyxovirus, contains a nonsegmented minus-strand RNA genome of 15.3 kilobases from which six mRNAs (NP, P/C, M, F, HN, and L) are transcribed. The polymerase responsible for mRNA synthesis is thought to be composed of the viral P and L proteins (15), and the template is the genome nucleocapsid. The nucleocapsid is a helical structure in which the RNA is tightly associated with about 2,000 copies of the NP protein and has some morphological similarities to tobacco mosaic virus. mRNA synthesis is thought to take place by the polymerase, beginning at the 3' end of the template and making a short 5' leader RNA and then each of the mRNAs in turn by a stop-start mechanism. The junctions between mRNAs on the genome contain conserved sequences, which are thought to specify polyadenylation-termination of each mRNA and reinitiation of the next. The mRNAs are capped as well as polyadenylated, and these modifications are also thought to be due to the viral polymerase (reviewed in reference 20).

Except for the mRNA of the P/C gene, the second gene in the transcriptional map, all of the mRNAs are monocistronic in that they give rise to a single primary translation product. The P/C-gene mRNA (Fig. 1), however, is remarkable in that it not only contains the overlapping P- and C-protein open reading frames (ORFs) (13, 33) but expresses more than one protein from at least the C ORF by ribosomal initiation at multiple start sites (7–9, 14, 25). The initiation codon closest to the 5' end of the mRNA is the ACG at position 81, which starts the C' protein (215 amino acids [aa]), and is followed by AUGs at positions 104 (P protein; 568 aa), 114 (C protein; 204 aa), 183 (Y1 protein; 181 aa), and 201 (Y2 protein; 175 aa). The C', C, Y1, and Y2 proteins, collectively referred to as the C proteins, are all in the −1 reading frame relative to P and represent a nested set of proteins with staggered N termini.

Besides Sendai virus, most paramyxovirus P genes code for a P protein of 500 to 600 aa, with a C ORF of about 200 aa overlapping the N terminus of P (e.g., human and bovine parainfluenza virus 3 [12, 21, 29, 34], measles virus [2], and

canine distemper virus [1]). In three other viruses (simian virus type 5 [SV5; 36], mumps virus [35], and Newcastle disease virus [30]), the P gene codes for a P protein of just under 400 aa, and the C ORF is missing. Nevertheless, even in these cases the P gene appears to code for more than one primary translation product (6, 16, 22, 23, 26). One manner in which this can take place has recently become clear for SV5 (36). The SV5 P gene codes for a V protein (ca. 200 aa) in addition to P, which, unlike the Sendai virus P and C proteins, share tryptic peptides. The SV5 P gene contains a stop codon in the middle, and the mRNA that is an exact copy of the genome can be translated into the V and not the P protein. The SV5 P protein was found to be translated from a second mRNA, which contained a precise two-G insertion upstream of the internal stop codon, allowing access at this point to the +1 ORF, which continued through the remainder of the mRNA. The C-terminal portion of V, which is not present in P, is unusually rich in Cys residues and was found to be well conserved among all paramyxoviruses (except respiratory syncytial virus), in sharp contrast to the P and C sequences. Thomas et al. (36) therefore predicted that a V-like protein would also be expressed from the other paramyxovirus P genes. Curiously, in contrast to SV5, in all of these other viruses the uninserted mRNA would code for the longer P protein.

Further evidence was not long in coming. Cattaneo et al. (4) independently found that the measles virus P gene also expressed two mRNAs, which varied only by a precise one-G insertion. The uninserted copy translated into the P and C proteins, and the inserted mRNA translated into a V protein as well as C. In both SV5 and measles virus, one or two G's were inserted within a short run of G's found within a similar sequence context. Since similar sequences were present upstream of the V ORF in the other paramyxoviruses, Cattaneo et al. (4) and Paterson et al. (24) predicted where the putative insertions would occur. For Sendai virus, this would be a one-G insertion following the GGG at position 1053 (Fig. 1).

The circumstantial evidence that a V protein would also be expressed in Sendai virus was compelling. However, al-
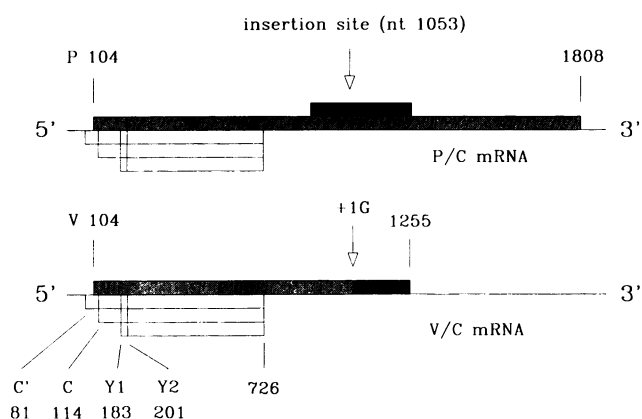
---

* Corresponding author.

FIG. 1. Protein-coding regions of the P/C and V/C mRNAs. The mRNA is shown as a horizontal line, with the 5' and 3' ends indicated. The P, C, and V ORFs are shown as hatched, open, and black boxes, respectively. The position of each start codon and the protein that results are indicated. Numbers refer to the first base of the start or stop codons relative to the 5' end of the mRNA. Note that the unaltered (P) and inserted (V) mRNAs would translate the C proteins identically.

though several primary translation products from the P gene were well documented, neither we nor others had ever noticed a V protein. The predicted protein would be 46 kilodaltons (kDa). However, if, like the P protein, it migrated anomalously slowly on gels, it might comigrate with the more abundant NP or Fo proteins. In this paper, we show that a Sendai virus P-gene mRNA with the predicted single-G insertion is present in infected cells and that this mRNA can be translated into a Cys-rich V protein. Moreover, the activity that modifies the mRNA by G insertion appears to be coded for by the virus and to act during mRNA synthesis.

MATERIALS AND METHODS

Cloning methods. Poly(A)$^+$ RNA (5 to 10 μg) or 0.25 μg of in vitro polymerase products containing 5 μg of uninfected-cell CsCl pellet RNA were mixed with 5 pmol of primer (the complement of either nucleotides [nt] 1863 to 1881 or nt 1155 to 1178) in 10 μl of ET (1 mM EDTA, 10 mM Tris chloride [pH 7.4]), boiled for 1 min, quickly chilled, and annealed for 10 min at 45°C in 0.2 M NaCl. First-strand synthesis was carried out in a total volume of 50 μl containing 50 mM Tris chloride (pH 8.3)–10 mM MgCl$_2$–30 mM 2-mercaptoethanol–0.5 mM deoxynucleoside triphosphates (dNTPs)–0.5 U of RNasin per μl–20 U of murine leukemia virus reverse transcriptase per μg of RNA. After 10 min at 37°C and 90 min at 42°C, the volume was raised to 250 μl with 0.25 mM β-NAD–86 mM KCl–10 mM ammonium sulfate–3 mM MgCl$_2$–10 mM Tris chloride (pH 7.5)–50 μg of bovine serum albumin–0.1 mM dNTPs–8.5 U of RNase H per ml–230 U of DNA polymerase I per ml–10 U of *Escherichia coli* DNA ligase per ml. After 60 min at 15°C and 60 min at 22°C, the reaction mixture was made 10 mM in EDTA and 0.2% in sodium dodecyl sulfate, phenol-chloroform extracted, and precipitated with 2 M ammonium acetate and 2 volumes of ethanol. The DNA was dissolved in ET, a one-third fraction was digested with restriction enzymes and reprecipitated as described above, and one-third of this was ligated with 30 ng of either pBluescript KS or SP65-P/C from which the *Xba*I-*Nde*I fragment (nt 1028 to 1716) was removed. Both vectors

were prepared by digestion with the appropriate restriction enzymes. The DNA was then separated on a low-melting-point agarose gel and purified on an Elutip-D (Schleicher & Schuell, Inc.) column. After ligation, a one-third fraction was used to transform competent *E. coli* DH5.

Either Nonidet P-40-disrupted purified virions or cytoplasmic extracts of infected BHK cells were centrifuged on 20 to 40% CsCl density gradients; the nucleocapsid band at 1.31 g/ml was harvested, and its RNA was isolated by phenol extraction. Approximately 0.5 μg of this RNA was mixed with 5 μg of uninfected-cell CsCl pellet RNA and cloned as described above except that the primer was complementary to the (minus-strand) genome (nt 983 to 1006 of the P gene).

When CsCl pellet RNA was used as the starting material, first-strand synthesis was carried out as described above except that 50 μg of RNA in a 300-μl reaction was used. The RNA-DNA hybrid was precipitated with ammonium acetate, and second-strand synthesis was carried in a 250-μl reaction as described above.

For synthesis of full-length clones, primers from both ends of the gene were used such that overhanging 5' ends on the resulting DNA could be generated for cloning (4). The primers used were (i) complementary to nt 1863 to 1881 of the P mRNA with an additional CCGG at its 5' end and (ii) the direct sequence of nt 7 to 26 with an additional CG at its 5' end.

Colony hybridization with oligos. The resulting colonies were replica streaked onto nitrocellulose filters. When the 5'-end-labeled genomic and +G oligonucleotides (oligos) were used, the filters were hybridized overnight at 38°C in 5× SSC (SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–5% skim milk (Difco Laboratories). For the neutral oligo, 33°C was used. The filters were washed twice for 5 min and once for 20 min in 1× SSC–0.1% sodium dodecyl sulfate at 22°C. For the genomic and +G oligos, a further wash of 20 min at 50°C was carried out.

RESULTS

A P mRNA in vivo with a precise G insertion. To determine whether P mRNAs found in vivo contained species with the predicted insertion, cDNA was prepared from infected-cell poly(A)$^+$ RNA by using a primer complementary to nt 1863 to 1881 of the mRNA (the 3' untranslated region). There are no unique restriction sites within the 5' untranslated region, and many of the cDNAs might not have extended this far. As incomplete clones would be of limited use for further translational studies but would nevertheless be scored by colony hybridization with probes to the insertion site, we cloned only a segment of the gene containing the proposed insertion site. The cDNA was digested with *Xba*I and *Nde*I (both unique; positions 1028 and 1716), and this mixture was cloned into a plasmid that contained a genomic copy of the gene (and whose mRNA translates into the P and C proteins; SP65-P/C) but from which this fragment had been removed. Replicates of the resulting colonies were then screened with three oligos (Fig. 2): two to the proposed site (genomic and +G oligos) and one to a sequence just downstream where there is no reason to believe that changes take place (neutral oligo). Of the 110 neutral-oligo-positive (neutral-oligo$^+$) clones examined, 13% were found to be +G oligo$^+$ and genomic oligo$^-$, whereas most of the remainder were found to be the reverse. Moreover, when mRNAs (via SP6 polymerase) from representatives of each group were translated in vitro (Fig. 3), those of the genomic-oligo$^+$ group made the P (and C) proteins as expected. Those of the +G-oligo$^+$

predicted insertion site
↓

A 5' GGUCUAGAGACCGAC UCAACAAAAAAGGGCAUAGG AGAGAACACAUCAUCUAUGAAAGAGAUGGC 3'

∗ (above the C in GACCGAC region)

B genomic   3' TGTTTTTTCCCGTATCCTC 5' 3' GTAGTAGATACTTTC 5' neutral

+G   3' GTTTTTTCCCGTATCCTC 5'
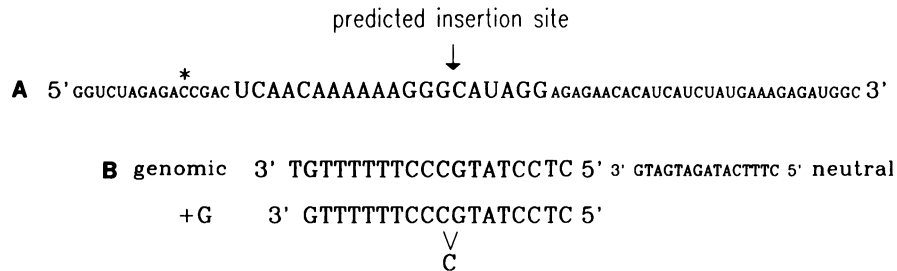   V
                              C

FIG. 2. Predicted insertion site and oligos used for screening. (A) Sequence of positions 1025 to 1089 of the mRNA. The region that bears homology to the other known and predicted insertion sites is in larger letters. The C marked with an asterisk (nt 1035) is a G in the gene contained in the VV recombinant. (B) Sequences of the three DNA oligos (genomic, +G, and neutral) used in the colony hybridizations are shown.

group made, in addition to the C proteins, a Cys-rich protein that migrated with an apparent molecular mass of 60 kDa (which we refer to as V) and that did in fact comigrate with the NP protein (not shown). The Sendai virus V protein would then initiate like P at the AUG at position 104 and contain the first 316 aa of P fused to 68 aa of the Cys-rich V ORF (Fig. 1). The predicted mass of the V protein is 46 kDa. However, the 62-kDa P protein also migrated anomalously on gels as an 80-kDa species. The determinant of this anomalous migration therefore appears to be located in the N-terminal regions of the P and V proteins, which are particularly acidic.

Frequency of G insertions at position 1053. The cloning protocol described above probably underestimates the frequency of mRNAs with G insertions. Some of the SP65-P/C DNA used might have been cut with only one of the two enzymes, leading to a background of genomic-oligo$^+$ colonies on self-ligation. The same cDNA was therefore digested

FIG. 3. In vitro translation of the cloned mRNAs. DNAs from three +G-oligo$^+$ clones and one genomic-oligo$^+$ clone (see text), as well as the original SP65-P/C plasmid, were transcribed with SP6 polymerase. The mRNAs were then translated in a rabbit reticulocyte lysate, the proteins were labeled with either [^{35}S]methionine (M) or [^{35}S]cysteine (C), and the products were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The protein products are indicated in the margins.

with XbaI and EcoRI (nt 1028 to 1131), and this mixture was cloned into virgin pBluescript. Replicate colonies were then screened as described above (Fig. 4). Of the 113 neutral-oligo$^+$ colonies examined, 31% were now +G oligo$^+$ (Table 1, line 1), and three of three clones of this group were found to contain a one-G insertion at nt 1053 by direct sequencing of the plasmid DNA (Fig. 5). In addition, four of four clones of the genomic-oligo$^+$ group contained no insertions at nt 1053. As a further control, five full-length clones were isolated from the same RNA (see Methods and Materials). The mRNAs from three of these clones translated into the P protein, and the mRNAs of two translated into the V protein (not shown). Thus, approximately a third of the P mRNAs in vivo contained this precise G insertion.

To determine that insertions within the P mRNA actually take place, it is necessary to demonstrate that this region of the viral genome is unique. Moreover, to exclude the possibility that virion genomes might somehow not be representative of intracellular genomes, the latter were also examined. Virion and intracellular nucleocapsids were first separated from mRNA by banding in CsCl density gradients, and the insertion region of the genome RNA was cloned as before except that the primer used was complementary to the negative-strand genome and upstream of the insertion site. Of the 136 virion and the 152 intracellular genome clones examined that were neutral oligo$^+$, 134 and 149, respectively, were found to be genomic oligo$^+$, and none were +G oligo$^+$. When the five remaining clones were sequenced, three did not contain viral sequences and two were found to contain the genomic sequence. Thus, no G insertions at nt 1053 could be found in a total of 285 clones (Table 1, lines 8 and 9). The vast majority, if not all, of the G insertions in the mRNA that we detected must therefore have occurred during or after mRNA synthesis.

The frequency of insertions appears to be fixed. We next sought to determine whether the frequency of G insertions was a fixed ratio or whether it varied during different viral infections. During strain Z infections, for example, the viral polymerase behaves somewhat differently than during strain H infections (examined above). The strain Z polymerase fails to terminate at the leader RNA stop site at a frequency which is 20 to 40 times higher than that during strain H infections, making leader-NP readthrough RNAs (37). Also, during coinfection with defective interfering particles, the viral polymerase is engaged predominantly in genome replication as opposed to mRNA synthesis (19), and the infection becomes persistent as opposed to lytic (28). To compare these situations, total unencapsidated RNA was isolated from parallel infections with strain H, strain H plus defective
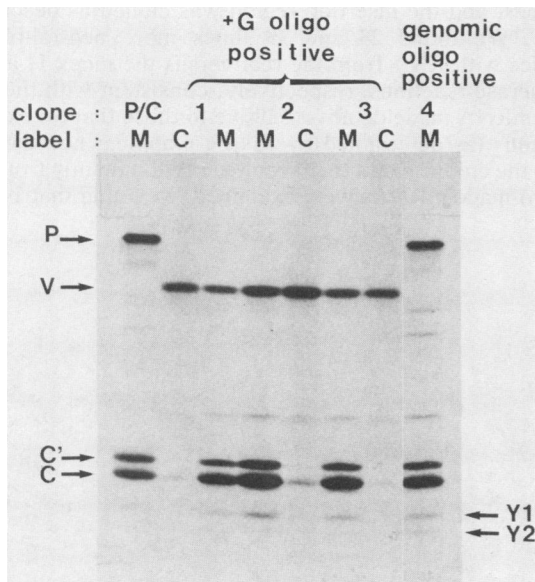
TABLE 1. Frequency of G insertions at position 1053

Line	Strain	RNA examined	No. of colonies with inserts	No. (%) of colonies with given no. of G's inserted			Total insertion frequency (%)
				0	1	>1	
1	H	Poly(A)$^+$	113	70 (62)	35 (31)	8 (7)	38
2	H	CsCl pellet	177	119 (68)	48 (27)	10 (6)	33
3	H/DI	CsCl pellet	184	108 (59)	62 (34)	14 (7)	41
4	Z	CsCl pellet	122	71 (58)	39 (32)	12 (10)	42
	Avg ± mean (%)			62 ± 3.2	31 ± 2	7.5 ± 1.2	38.5 ± 3.0
5	H	In vitro	118	92 (78)	17 (14)	9 (8)	22
6	Z	In vitro	245	187 (76)	49 (20)	9 (4)	24
7	Z	In vitro	154	112 (73)	28 (18)	14 (9)	27
	Avg ± mean (%)			76 ± 1.7	17 ± 2.2	7.0 ± 2	24.3 ± 1.5
8	Z	Virion CsCl band	135	135 (100)	0	0	0
9	Z	Intracellular CsCl band	150	150 (100)	0	0	0
10	VV-P/C	5.7 M CsCl pellet	198	198 (100)	0	0	0
11	Z + VV-P/C	Sendai virus mRNA	229	148 (64)	72 (32)	9 (4)	36
		VV mRNA (5.7 M CsCl pellet)	150	150 (100)	0	0	0

interfering particles (H4), and strain Z and examined as described above. We found very little difference in the frequency of single G insertions, which varied only from 27 to 34% (Table 1, lines 2 to 4). Furthermore, two separate strain H infections gave frequencies of 27 and 31% (lines 1 and 2). When these four infections were treated as a group, we found that the mRNA was unaltered at this site 62 ± 4% of the time, a single G was inserted 31 ± 2% of the time, and 7.5 ± 1% of the clones could not be accounted for by these two events (see below). The frequency with which the G insertion was made therefore did not appear to vary during these different infections of the same BHK cells.

A virus-encoded activity inserts the G residues. The insertions have been suggested to be due to a very precise stuttering, or reiterative copying, by the viral polymerase during transcription of the short run of C's on the template (4, 36). Polyadenylation of mRNAs is thought to take place by a similar but much less precise mechanism at the short run of U's at each mRNA stop site. However, given the very constant frequency of insertion among the different types of Sendai virus infection noted above, we would like to know

whether the insertions are in fact due to virus-encoded proteins and whether cellular factors are also involved. Moreover, there is no evidence so far which excludes the possibility that the insertions are due to some form of editing that is nonviral. This question can be addressed with purified Sendai virions, which make significant amounts of mRNAs in vitro.

Virion polymerase reactions of both strains H and Z were carried out, and parallel mock reactions containing no triphosphates and excess EDTA were also performed (37). The mock reactions are required to assess the level of preexisting mRNAs that might be present as virion contaminants. The mRNAs were then separated from the virion genomes, and the insertion region was cloned as described above. We found 23 and 39 times more neutral-oligo$^+$ colonies with RNA from the real versus the mock H and Z polymerase reactions, respectively, consistent with the levels of mRNA made de novo relative to those that preexisted in virions (as determined by RNase mapping; not shown). When the colonies that then resulted predominantly from the in vitro-made mRNAs were examined, we found that 14 and
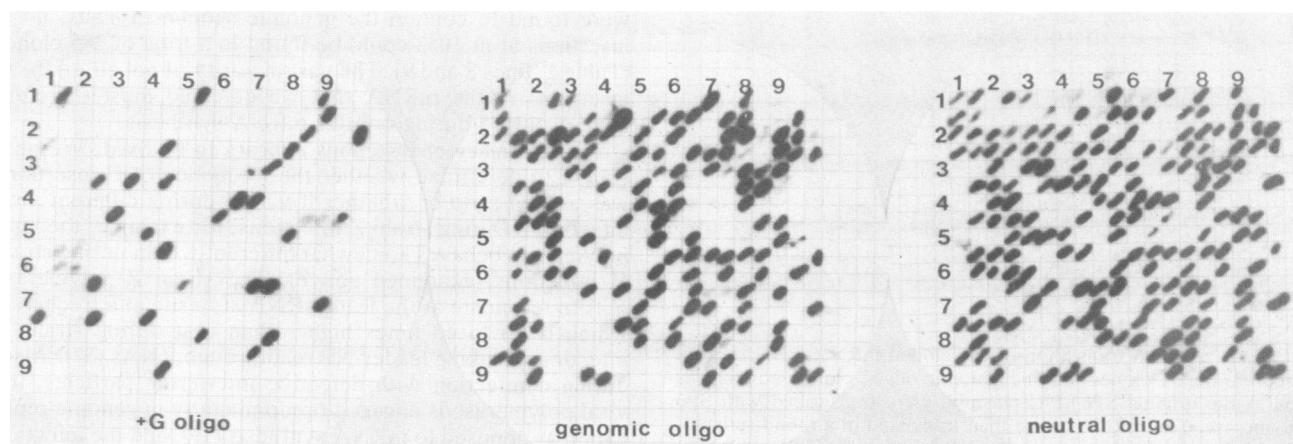

FIG. 4. Detection of insertions at position 1053 by colony hybridization. A fixed number of colonies (323) that resulted from cloning of the XbaI-EcoRI fragment of the P mRNA and that would fill a 13-cm filter were replica streaked onto four filters. Three of these filters were hybridized with either the neutral, genomic, or +G oligo.
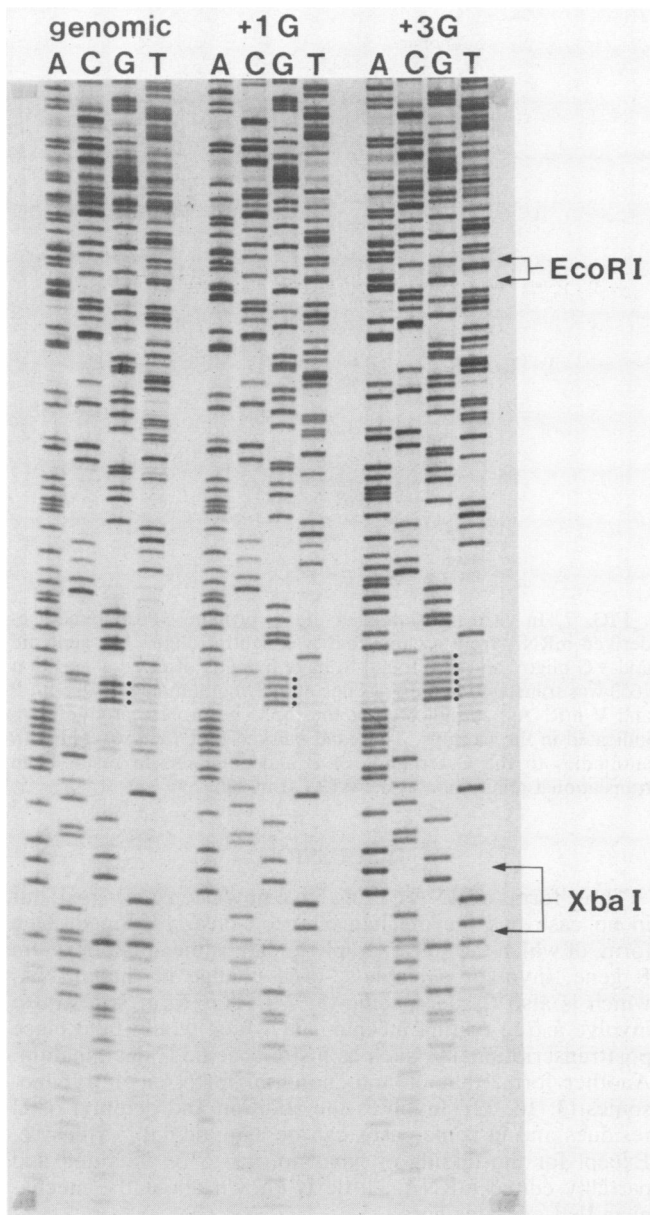
FIG. 5. DNA sequencing of the mRNA clones. DNAs from one genomic and one +G-oligo$^+$ clone and from one neutral-oligo$^+$ clone that was both genomic and +G oligo$^-$ were sequenced by the dideoxy-NTP method. The G's at position 1053 are marked with dots to the right of the G lanes. The XbaI and EcoRI sites used in the cloning are indicated on the right.

18 to 20% of the H and Z virion products, respectively, were +G oligo$^+$ (Table 1, lines 5 to 7). Representatives of each group were again confirmed by plasmid DNA sequencing.

The insertions were therefore likely to be the result of a virus-encoded activity, but there is a caveat to this conclusion. Cellular actin is visible in stained protein gels of our purified virions, and host tRNAs (17), as well as RNase NU and RNase P (18), can be detected enzymatically. It is therefore not impossible that a cellular editing activity has been similarly included in virions. If this were so, however, we would expect that this activity would also edit the P/C mRNA expressed from a recombinant vaccinia virus (VV)

carrying a genomic copy of this gene (VV-P/C). Cytoplasmic RNA was therefore prepared from BHK cells infected with VV-P/C when P and C protein expression was maximal (9) by sedimentation through a cushion of 5.7 M CsCl to remove DNA. We controlled the absence of VV DNA from our sample by showing that no neutral-oligo$^+$ colonies could be generated by cloning without prior cDNA synthesis. The insertion region of the recombinant mRNA was then cloned and examined. All of the 198 neutral-oligo$^+$ clones examined were also genomic oligo$^+$ (Table 1). We found only three ambiguous clones, which were also +G oligo$^+$ but which on sequencing had three G's at nt 1053. The P/C mRNA expressed from VV DNA was therefore not modified in the same cell line in which the mRNA expressed from Sendai virus was modified by a single G insertion 31% of the time.

If the insertions resulted from the polymerase stuttering during transcription, then the insertion activity would not operate on preformed mRNAs. This could be tested by examining the VV-derived P/C mRNA from cells that were coinfected with Sendai virus if the natural and the VV mRNAs could be distinguished. Serendipitously, we found a single-base change within the XbaI-to-EcoRI region between VV-P/C that was constructed from a clone obtained in 1981 (a G at nt 1035) and our current Sendai H virus stocks (a C at nt 1035). We assume that this base difference does not affect the frequency of insertion at nt 1053, since this virus strain has had the same growth characteristics in different cell lines over these years. Oligos that could distinguish a G or a C at nt 1035 (3'-GAGATCTCT[G/C]GCTGAGTTG-5') were then made and used to first separate cDNA clones of mRNA from coinfected cells according to which viral genome they had been transcribed from. These clones were then examined for G insertion at nt 1053. We found that none of the VV mRNAs were altered during the same infection in which 36% of the natural Sendai virus mRNA had insertions (Table 1, line 11). Since the coinfection was started at the same time and there was five times less VV mRNA than the natural mRNA, as judged by the number of clones of each type, it seems unlikely that the lack of insertions in the VV mRNA was due to limiting insertion activity. It therefore appears that the insertion activity acted only during RNA synthesis by the Sendai virus polymerase. This experiment also ruled out the possibility that the activity was host encoded but induced only by Sendai virus infection.

Together, these results indicate that the activity responsible for the insertions is virus encoded and cannot function in *trans*. However, the somewhat lower insertion frequency in vitro could mean that cellular proteins help in this process. Alternatively, it might reflect other differences between the in vitro and in vivo conditions for mRNA synthesis.

Misinsertion at position 1053. In the various RNA samples examined (Table 1), 90% or more of the neutral-oligo$^+$ colonies were either genomic oligo$^+$ or +G oligo$^+$. However, 4 to 10% of these colonies were neither, and they appeared too frequently to be explained simply by the level of false-positives and -negatives during colony hybridization (1 to 2%; see above). It appeared that insertions other than a single G were taking place but at a lower frequency. In the case of measles virus, of the 19 clones examined, 7 contained one G and two contained three G's inserted at this site. Moreover, in a separate study, a clone with 10 G's at this site had been noted (4).

We therefore sequenced this region of the plasmid DNA from 12 neutral-oligo$^+$ colonies that were neither genomic nor +G oligo$^+$, obtained from both in vivo and in vitro-made mRNAs. We found that in contrast to the genomic-oligo$^+$
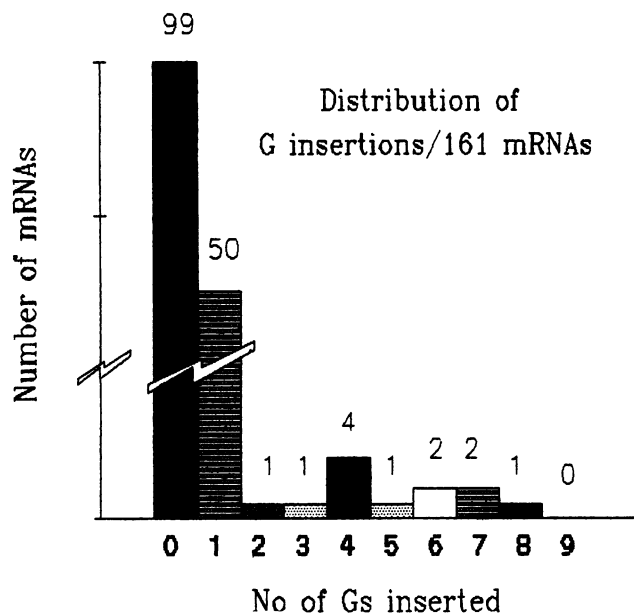
FIG. 6. Normalized distribution of mRNAs according to the number of G's inserted. The distribution of G insertions in 12 clones with more than one G inserted at position 1053 is shown. These clones represent 7.5% of the population, and the normalized numbers of mRNAs with no or one G inserted are derived from the data in Table 1.

FIG. 7. In vitro translation of the W protein. SP6 polymerase-derived mRNA from a clone that was neutral oligo$^+$ but genomic and +G oligo$^-$ and was found to have five G's inserted at position 1053 was translated in the presence of [^{35}S]methionine, along with P and V mRNAs, as indicated at the top. The protein products are indicated in the margins. The band marked Met 1369 reacted with antibodies to the C terminus of P and appeared to result from translational reinitiation at the AUG at nt 1369.

plasmids that contain three G's at this site or the +G-oligo$^+$ plasmids that contain four G's, these 12 plasmids contained 5 to 11 G's, the distribution of which is given in Fig. 6. No other sequence changes were found at any other site between positions 1028 and 1131 among a total of 28 clones examined. Thus, similar to the case with measles virus, Sendai virus P mRNAs that were altered from the genomic sequence contained predominantly a single G insertion (27 to 34%). However, 4 to 10% of these mRNAs contained multiple (two to eight) G insertions at this site. Since viral genomes with only three G's at nt 1053 are present in our infections, then multiple as well as single G insertions must occur during or after mRNA synthesis.

P mRNAs with a single G insertion at nt 1053 shift into the −1 ORF relative to that which translates P. However, mRNAs with two, five, or eight G's inserted would shift into the +1 ORF, which contains a stop codon two codons downstream. These mRNAs should therefore make a truncated form of P with only the last 2 aa from an alternate ORF. In vitro translation of these mRNAs showed that this was indeed so (Fig. 7). We refer to this truncated form of P as W and expect that it will also be present in infected cells but at a much lower level than P or V. Since there is little bias for the precise number of insertions when more than one G is added, we expect one-third of the 7.5% multiple-G-insertion mRNAs to code for W. Assuming that the P, V, and W proteins are equally stable, infected cells would contain twice as much P as V and roughly 12 times as much V as W, as determined from the results in Table 1. We note, however, that neither V nor W has as yet been detected in vivo and presume that this is due to the absence so far of suitable antibodies. We have screened a panel of 25 anti-P monoclonal antibodies from Claes Orvell (Karolinska Institute, Stockholm, Sweden), but all were found to react with epitopes downstream of the insertion site.
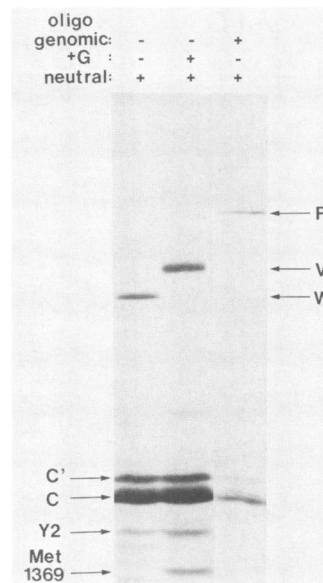
DISCUSSION

Three forms of RNA editing have now been described, but in no case has the mechanism been clearly defined. One form, of which the sole example to date is the apolipoprotein B gene, involves a single C-to-U change in the mRNA, which is also tissue specific (5, 27). This form is likely to involve a true editing mechanism, which would take place posttranscriptionally, such as a site-specific C deamination. Another form, that of mitochondrial mRNAs of trypanosomes (3, 10, 32), involves the addition and deletion of U residues and in some cases can be astoundingly extensive. Except for the finding of what appear to be unedited and partially edited mRNAs, little is known about the mechanism that operates here.

The most recently described form is that of the paramyxovirus P-gene mRNAs, in which either one or two G's are added in the middle of an ORF to gain access at this point to an alternate reading frame. In all three viruses examined, the insertions take place at a precise site within a run of three to four G's. For SV5, two G's are inserted to shift into the C-terminal half of the P reading frame. For measles and Sendai virus, a single G is added, and this in contrast results in a shift to the V ORF. Here the insertion mechanism is not absolutely specific. Two to eight G's are added at a total frequency of 6 to 10% in vivo, under conditions in which one G is added 27 to 34% of the time. For SV5, no misinsertions were detected; of the 22 clones reported, 10 contained two extra G's and 12 were unaltered. However, in light of our data, this number of clones may have been insufficient to detect misinsertion on statistical grounds.

We have shown that the single- and multiple-base insertions most likely occur during mRNA synthesis. When multiple insertions occur, although there is little bias for the exact number of bases inserted (two to eight), in every case

only G's are inserted. Moreover, in the numerous examples of misediting detected so far, no examples of deletions have been found. For Sendai virus, the insertion of single and multiple G's also takes place in vitro, using purified virions as the source of both polymerase and template. Moreover, when the P/C mRNA is expressed in vivo via a VV recombinant in either the presence or absence of the other Sendai virus proteins, no insertions occur. This finding suggests that virus-encoded proteins are essential for this process and that they cannot act in *trans* on the VV-derived mRNA, presumably because the insertions take place cotranscriptionally. However, we cannot rule out the possibility that the VV mRNAs are unaltered because of the extra sequences present at both ends of these mRNAs, even though the central 1,800 nt are the same as in the natural mRNA. Only analysis of virus mutants with altered insertion activity can settle this point unambiguously.

All of these findings are consistent with the stuttering mechanism suggested previously, in which the viral polymerase reiteratively copies while crossing the three to four C's on the template. Such a mechanism is also attractive in that it is similar to what is thought to happen during polyadenylation of these mRNAs but at a run of five to seven U's. There are, of course, notable differences between the insertions and polyadenylation. Polyadenylation is coupled to termination, and the viral polymerase almost always polyadenylates at this site during mRNA synthesis. For Sendai virus, this fails to occur only 0.5% of the time in vivo (37). Stuttering would also occur for a much longer time here, and there is little requirement for the number of stutters to be precise (100 to 400 A's are added). The basic mechanism nevertheless remains the same, and in neither case would the polymerase stutter at all during genome replication. Stuttering has also recently been suggested as the mechanism by which poly(A) leader formation take place on the late VV mRNAs (31).

Within a given host cell in culture, the frequency of G insertions does not vary among different Sendai virus infections but appears to be a relatively fixed ratio. In terms of a stuttering mechanism, the polymerase would not stutter 62% of the time and would stutter precisely once 31% of the time. In such a facultative mechanism, we would have expected that if the polymerase were to stutter more than once, there would have been a preference for twice, then three times, etc. However, this does not seem to be so for either Sendai or measles virus, and it is difficult to see how the SV5 polymerase would stutter precisely twice when it added extra G's. We also note that none of the information we now have concerning the insertions is sufficient to exclude the possibility that they occur after the polymerase has crossed this region. The in vitro system we describe may, however, help in determining whether the stuttering mechanism is correct.

Inspection of the known and predicted insertion sites among paramyxoviruses (4, 24) shows a weak but clear consensus sequence 5'-AARRGGG CAXRG (the space indicates the insertion site), but even this sequence is not uniformly conserved. This finding suggests that this sequence plays only a limited role in determining the insertion site and offers no clues as to why one or two G's are inserted in the different viruses. Moreover, except for the five to seven U's on which polyadenylation occurs, there is even less conservation of flanking sequences among paramyxoviruses at this site. It therefore seems reasonable that determinants other than linear sequence play a role in specifying these exact sites. For SV5, Thomas et al. (36) noted that the

insertion site could theoretically be found as a loop within a stem-loop structure in the template, but similar structures are unlikely for either Sendai or measles virus. Thomas et al. (36) also noted, however, that spurious bands appeared during dideoxy-NTP sequencing of the insertion region of genomic clones, which were absent from clones which had insertions. Curiously, the same effect (but of lesser magnitude) can be seen with Sendai virus in Fig. 5, where there is a band in all four lanes at nt 1053 during sequencing of the genomic but not the inserted clones. One interpretation of these results is that the Klenow and T7 polymerases stop or pause at this site in genomic clones because they sense a structural constraint, which is alleviated by the G insertions. It is unclear whether these effects on DNA can be related to the RNA in paramyxovirus nucleocapsids. Nevertheless, RNA structures (either within the template or the nascent mRNA chain or possibly even interactions of the nascent chain and the template outside of the insertion site) would appear to be required both to account for the specificity of the site and to explain how one or two G's are added by the different paramyxoviruses.

LITERATURE CITED

1. **Barret, T., S. B. Shrimpton, and S. E. H. Russel.** 1985. Nucleotide sequence of the entire protein coding region of canine distemper virus polymerase-associated (P) protein mRNA. Virus Res. 3:367–372.
2. **Bellini, W. J., G. Englund, S. Rozenblatt, H. Arnheiter, and C. D. Richardson.** 1985. Measles virus P gene codes for two proteins. J. Virol. 53:908–919.
3. **Benne, R., J. Van Den Burg, J. P. J. Brakenhoff, P. Sloof, J. H. Van Boom, and M. C. Tromp.** 1986. Major transcript of the frameshifted coxII gene from Trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46:819–826.
4. **Cattaneo, R., K. Kaelin, K. Baczko, and M. A. Billeter.** 1989. Measles virus editing provides an additional cysteine-rich protein. Cell 56:759–764.
5. **Chen, S. H., G. Habib, C.-Y. Yang, Z.-W. Gu, B. R. Lee, S. A. Weng, S. R. Silberman, S. J. Cai, J. P. Deslypere, M. Rosseneu, A. M. Gotto, Jr., W.-H. Li, and L. Chan.** 1987. Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. Science 238:363–366.
6. **Collins, P. L., G. W. Wertz, L. A. Ball, and L. E. Hightower.** 1982. Coding assignments of the five smaller mRNAs of Newcastle disease virus. J. Virol. 43:1024–1031.
7. **Curran, J. A., and D. Kolakofsky.** 1988. Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. EMBO J. 7:245–251.
8. **Curran, J. A., and D. Kolakofsky.** 1988. Scanning independent ribosomal initiation of the Sendai virus X protein. EMBO J. 7:2869–2874.
9. **Curran, J. A., and D. Kolakofsky.** 1989. Scanning independent ribosomal initiation of the Sendai virus Y proteins in vitro and in vivo. EMBO J. 8:521–526.
10. **Feagin, J. E., J. M. Abraham, and K. Stuart.** 1988. Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei. Cell 53:413–422.
11. **Feagin, J. E., D. P. Jasmer, and K. Stuart.** 1987. Developmentally regulated addition of nucleotides within apocytochrome b transcripts in Trypanosoma brucei. Cell 49:337–345.
12. **Galinski, M. S., M. A. Mink, D. M. Lambert, S. L. Wechsler, and M. W. Pons.** 1986. Molecular cloning and sequence analysis of the human parainfluenza 3 virus mRNA encoding the P and C proteins. Virology 155:46–60.
13. **Giorgi, C., B. M. Blumberg, and D. Kolakofsky.** 1983. Sendai virus contains overlapping genes expressed from a single mRNA. Cell 35:829–836.
14. **Gupta, K. C., and S. Patwardhan.** 1988. ACG, the initiator codon for a Sendai virus protein. J. Biol. Chem. 263:8553–8556.
15. **Hamaguchi, M., T. Yoshida, K. Nishikawa, H. Naruse, and Y.**

Nagai. 1983. Transcriptive complex of Newcastle disease virus. I. Both L and P proteins are required to constitute an active complex. Virology 128:105–117.

16. Herrler, G., and R. W. Compans. 1982. Synthesis of mumps virus polypeptides in infected vero cells. Virology 100:433–449.

17. Kolakofsky, D. 1972. tRNA nucleotidyltransferase and tRNA in Sendai virions. J. Virol. 10:559–563.

18. Kolakofsky, D., and S. Altman. 1978. Endoribonuclease activity associated with animal RNA viruses. J. Virol. 25:274–284.

19. Kolakofsky, D., and B. M. Blumberg. 1982. A model for the control of non-segmented negative strand viruses genome replication, p. 203–212. In B. W. J. Mahy, A. C. Minson, and G. K. Darby (ed.), Virus persistence symposium 33. Cambridge University Press, Cambridge.

20. Kolakofsky, D., and L. Roux. 1987. The molecular biology of paramyxoviruses, p. 277–297. In P. Bercoff (ed.), The molecular basis of viral replication. Plenum Publishing Corp., New York.

21. Luk, D., A. Sanchez, and A. K. Banerjee. 1986. Messenger RNA encoding the phosphoprotein (P) gene of human parainfluenza virus 3 is bicistronic. Virology 153:318–325.

22. McGinnes, L., C. McQuain, and T. Morrison. 1988. The P protein and the non-structural 38K and 29K proteins of Newcastle disease virus are derived from the same open reading frame. Virology 164:256–264.

23. Paterson, R. G., T. J. R. Harris, and R. A. Lamb. 1984. Analysis and gene assignment of mRNAs of a paramyxovirus, simian virus 5. Virology 138:310–323.

24. Paterson, R. G., S. M. Thomas, and R. A. Lamb. 1989. Specific non-templated nucleotide addition to a simian virus 5 mRNA: prediction of a common mechanism by which unrecognized hybrid P-cysteine-rich proteins are encoded by paramyxovirus P genes, p. 232–245. In D. Kolakofsky and B. W. J. Mahy (ed.), Genetics and pathogenicity of negative stranded viruses. Elsevier/North-Holland Publishing Co., Amsterdam.

25. Patwardhan, S., and K. C. Gupta. 1988. Translation initiation potential of the 5' proximal AUGs of the polycistronic P/C mRNA of Sendai virus: a multipurpose vector for site specific mutagenesis. J. Biol. Chem. 263:4907–4913.

26. Peluso, R. W., R. A. Lamb, and P. W. Choppin. 1977. Polypeptide synthesis in simian virus 5-infected cells. J. Virol. 23:177–187.

27. Powell, L. M., S. C. Wallis, R. J. Pease, Y. H. Edwards, T. J. Knott, and J. Scott. 1987. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell 50:831–840.

28. Roux, L., and J. J. Holland. 1979. Role of defective interfering particles of Sendai virus in persistent infections. Virology 93:91–103.

29. Sakai, Y., S. Suzu, T. Shioda, and H. Shibuta. 1987. Nucleotide sequence of the bovine parainfluenza 3 virus genome: its 3' end and the gene of NP, P, C and M proteins. Nucleic Acids Res. 15:2927–2944.

30. Sato, H., M. Oh-Hira, N. Ishida, Y. Imamura, S. Hattori, and M. Kawakita. 1987. Molecular cloning and nucleotide sequence of P, M, and F genes of Newcastle disease virus avirulent strain D26. Virus Res. 7:241–255.

31. Schwer, B., and H. G. Stunnenberg. 1988. Vaccinia virus late transcripts generated in vitro have a poly(A) head. EMBO J. 7:1183–1190.

32. Shaw, J. M., J. E. Feagin, K. Stuart, and L. Simpson. 1988. Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. Cell 53:401–411.

33. Shioda, T., Y. Hidaka, T. Kanda, H. Shibuta, A. Nomoto, and K. Iwasaki. 1983. Sequence of 3,687 nucleotides from the 3' end of Sendai virus genome RNA and the predicted amino acid sequences of viral NP, P and C protein. Nucleic Acids Res. 11:7317–7330.

34. Spriggs, M. K., and P. L. Collins. 1986. Sequence analysis of the P and C protein genes of human parainfluenza virus type 3: patterns of amino acid sequence homology among paramyxovirus proteins. J. Gen. Virol. 67:2705–2719.

35. Takeuchi, K., M. Hishiyama, A. Yamada, and A. Sugiura. 1988. Molecular cloning and sequence analysis of mumps virus gene encoding the P protein: mumps virus P gene is monocistronic. J. Gen. Virol. 69:2043–2049.

36. Thomas, S. M., R. A. Lamb, and R. G. Paterson. 1988. Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. Cell 54:891–902.

37. Vidal, S., and D. Kolakofsky. 1989. Modified model for the switch from Sendai virus transcription to replication. J. Virol. 63:1951–1958.