Two mRNAs That Differ by Two Nontemplated Nucleotides Encode the Amino Coterminal Proteins P and V of the Paramyxovirus SV5

Sheila M. Thomas, Robert A. Lamb,
and Reay G. Paterson
Department of Biochemistry, Molecular Biology,
and Cell Biology
Northwestern University
Evanston, Illinois 60208

Summary

The "P" gene of the paramyxovirus SV5 encodes two known proteins, P ($M_r \approx 44{,}000$) and V ($M_r \approx 24{,}000$). The complete nucleotide sequence of the "P" gene has been obtained and is found to contain two open reading frames, neither of which is large enough to encode the P protein. We have shown that the P and V proteins are translated from two mRNAs that differ by the presence of two nontemplated G residues in the P mRNA. These two additional nucleotides convert the two open reading frames to one of 392 amino acids. The P and V proteins are amino coterminal and have 164 amino acids in common. The unique C terminus of V consists of a cysteine-rich region that resembles a cysteine-rich metal binding domain. An open reading frame that contains this cysteine-rich region exists in all other paramyxovirus "P" gene sequences examined, which suggests that it may have important biological significance.

Introduction

In recent years the catalog of mechanisms identified as having a role in the processing or modification of the initial RNA transcripts to yield mature mRNAs has increased markedly. In eukaryotic cells the most common process found is splicing of the primary RNA transcript (Padgett et al., 1986). Less common variations on this splicing theme are alternative splicing, such that exons are excluded from some, but not all, of the mature mRNA, giving rise to sequence diversity in the encoded proteins (for review, see Breitbart et al., 1987), and *trans* splicing, in which two independently transcribed RNAs are ligated to form the mature mRNA (Konarska et al., 1985; Solnick, 1985; Murphy et al., 1986; Sutton and Boothroyd, 1986; Krause and Hirsh, 1987; Koller et al., 1987). Mechanisms of RNA transcript modification other than splicing include the phenomenon of RNA-editing in mitochondrial transcripts from trypanosomes, which is characterized by the presence in the mature mRNA of uridine residues that are not encoded in the gene (Benne et al., 1986; Feagin et al., 1987, 1988; Shaw et al., 1988). In addition, a process related to RNA-editing is thought to occur in primary transcripts of the mammalian apolipoprotein-B gene, as two discrete mRNAs have been found, one of which has a U residue in place of a templated C (Powell et al., 1987; Chen et al., 1987).

In animal virus infected cells, many examples of spliced

and alternatively spliced viral mRNAs have been identified. In addition, an unusual cotranslational modification has been identified in vaccinia virus late transcripts that possess a poly(A) leader at the 5′ end not encoded by the virus genome (Bertholet et al., 1987; Schwer et al., 1987). The 5′ poly(A) region is thought to be added by the vaccinia polymerase "stuttering" at a series of T residues on the template DNA (Schwer and Stunnenberg, 1988).

In many virus systems, in addition to modification of the primary RNA transcripts, the maximum protein coding potential on an mRNA is exploited by the use of alternative translation strategies that also have a potential role in the regulation of viral gene expression. Such mechanisms include translation from overlapping reading frames as observed for the coat, lysis, and synthetase proteins of bacteriophage MS2 (Atkins et al., 1979) and the adenovirus E1B proteins (Bos et al., 1981); ribosomal frameshifting that occurs to yield the gag–pol fusion proteins of Rous sarcoma virus, human immunodeficiency virus, or mouse mammary tumour virus (Jacks and Varmus, 1985; Jacks et al., 1987; Varmus, 1988), and that also occurs in the polymerase encoding region of infectious bronchitis virus (Boursnell et al., 1987; Brierley et al., 1987); and the use of suppressor tRNAs to overcome translation termination during translation of the gag–pol fusion protein in Moloney murine leukemia virus (Yoshinaka et al., 1985) and the nsP4 protein of Sindbis virus (Strauss et al., 1983).

In negative strand RNA viruses, several of the processes discussed above are involved in the regulation of viral gene expression. For instance, in influenza A and B viruses, both spliced and unspliced mRNAs that are translated to yield polypeptides from overlapping reading frames have been identified (Lamb and Lai, 1980; Briedis and Lamb, 1982), and influenza A viruses also provide an example of alternatively spliced mRNAs (Lamb et al., 1981). Translation from overlapping reading frames on functionally bicistronic mRNAs has been shown to be the mechanism used to yield the NA and NB glycoproteins of influenza B virus (Shaw et al., 1983) and the P and C proteins of the paramyxoviruses Sendai virus and parainfluenza virus 3 and the morbillivirus, measles virus (Giorgi et al., 1983; Gupta and Kingsbury, 1985; Curran et al., 1986; Luk et al., 1986; Galinski et al., 1986; Spriggs and Collins, 1986; Bellini et al., 1985).

Simian virus 5 (SV5), a prototype paramyxovirus, has a single-stranded, negative sense genomic RNA (vRNA) approximately 15,000 nucleotides in chain length that is transcribed in infected cells by the virion-associated RNA transcriptase to yield virus-specific mRNAs. The SV5 "P" gene has been shown to encode both the P protein ($M_r \approx 44{,}000$), and protein V ($M_r \approx 24{,}000$) by the arrest of translation in vitro of both P and V using a cDNA clone derived from SV5-specific mRNAs (Paterson et al., 1984). In addition, the P and V proteins have been shown to have tryptic peptides in common, although no precursor-product kinetics could be demonstrated (Paterson et al., 1984; our unpublished data). Protein P is a virion structural protein

```
.1                                                              .61          .80              .100
TGG CCG GAC GGG TTA GCA ACA AGC GAC TGC CGG TGC CAA CAG CGC AAT CCA CAA TCC ACA ATG GAT CCC ACT GAT CTG AGC TTC TCC CCA GAT GAG ATC AAT AAG
0                                               N-Terminus P and V   Met-Asp-Pro-Thr-Asp-Leu-Ser-Phe-Ser-Pro-Asp-Glu-Ile-Asn-Lys-(15)

          .120            .140            .160            .180            .200
CTC ATA GAG ACA GGC CTG AAT ACT GTA GAG TAT TTT ACT TCC CAA CAA GTC ACA GGA ACA TCC TCT CTT GGA AAG AAT ACA ATA CCA CCA GGG GTC ACA GGA CTA
0  Leu-Ile-Glu-Thr-Gly-Leu-Asn-Thr-Val-Glu-Tyr-Phe-Thr-Ser-Gln-Gln-Val-Thr-Gly-Thr-Ser-Ser-Leu-Gly-Lys-Asn-Thr-Ile-Pro-Pro-Gly-Val-Thr-Gly-Leu-(50)

          .220            .240            .260            .280            .300
CTA ACC AAT GCT GCA GAG GCA AAG ATC CAA GAG TCA ACT AAC CAT CAG AAG GGC TCA GTT GGT GGG GGT GCA AAA CCA AAG AAA CCG CGA CCA AAA ATT GCC ATT
0  Leu Thr Asn Ala Ala Glu Ala Lys Ile Gln Glu Ser Thr Asn-His-Gln-Lys-Gly-Ser-Val-Gly-Gly-Gly-Ala-Lys-Pro-Lys-Lys-Pro-Arg-Pro-Lys-Ile-Ala-Ile-(85)

          .320            .340            .360            .380            .400            .420
GTG CCA GCA GAT GAC AAA ACA GTG CCC GGA AAG CCG ATC CCA AAC CCT CTA TTA GGT CTG GAC TCC ACC CCG AGC ACC CAA ACT GTG CTT GAT CTA AGT GGG AAA
0  Val-Pro-Ala-Asp-Asp-Lys-Thr-Val-Pro-Gly-Lys-Pro-Ile-Pro-Asn-Pro-Leu-Leu-Gly-Leu-Asp-Ser-Thr-Pro-Ser-Thr-Gln-Thr-Val-Leu-Asp-Leu-Ser-Gly-Lys-(120)

              .440            .460            .480            .500            .520
ACA TTA CCA TCA GGA TCC TAT AAG GGG GTT AAG CTT GCG AAA TTT GGA AAA GAA AAT CTG ATG ACA CGG TTC ATC GAG GAA CCC AGA GAG AAT CCT ATC GCA ACC
0  Thr-Leu-Pro-Ser-Gly-Ser-Tyr-Lys-Gly-Val-Lys-Leu-Ala-Lys-Phe-Gly-Lys-Glu-Asn-Leu-Met-Thr-Arg-Phe-Ile-Glu-Glu-Pro-Arg-Glu-Asn-Pro-Ile-Ala-Thr-(155)
+1                                                                  His-Gly-Ser-Ser-Arg-Asn-Pro-Glu-Arg-Ile-Leu-Ser-Gln-Pro-
              .540            .560            .580            .600            .620
AGT TCC CCC ATC GAT TTT AAG AGG GGC AGG GAT ACC GGC GGG TTC CAT AGA AGG GAG TAC TCA ATC GGA TGG GTG GGA GAT GAA GTC AAG GTC ACT GAG TGG TGC
0  Ser-Ser-Pro-Ile-Asp-Phe-Lys-Arg-Gly-Arg-Asp-Thr-Gly-Gly-Phe-His-Arg-Arg-Glu-Tyr-Ser-Ile-Gly-Trp-Val-Gly-Asp-Glu-Val-Lys-Val-Thr-Glu-Trp-Cys-(190)
+1  -Val-Pro-Pro-Ser-Ile-Leu-Arg-Gly-Ala-Gly-Ile-Pro-Ala-Gly-Ser-Ile-Glu-Gly-Ser-Thr-Gln-Ser-Asp-Gly-Trp-Glu-Met-Lys-Ser-Arg-Ser-Leu-Ser-Gly-Ala-(191)

          .640            .660            .680            .700            .720
AAT CCA TCC TGT TCT CCA ATC ACC GCT GCA GCA AGG CGA TTT GAA TGC ACT TGT CAC CAG TGT CCA GTC ACT TGC TCT GAA TGT GAA CGA GAT ACT T AAT ACA GTG
0  Asn-Pro-Ser-Cys-Ser-Pro-Ile-Thr-Ala-Ala-Ala-Arg-Arg-Phe-Glu-Cys-Thr-Cys-His-Gln-Cys-Pro-Val-Thr-Cys-Ser-Glu-Cys-Glu-Arg-Asp-Thr. (222) C-Terminus V
+1  -Ile-His-Pro-Val-Leu-Gln-Ser-Pro-Leu-Gln-Gln-Gly-Asp-Leu-Asn-Ala-Leu-Val-Thr-Ser-Val-Gln-Ser-Leu-Ala-Leu-Asn-Val-Asn-Glu-Ile-Leu- Asn-Thr-Val-(226)

          .740            .760            .780            .800            .820            .840
AGA AAT TTG GAC TCT CGG ATG AAT CAA CTG GAG ACA AAA GTA GAT CGC ATT CTC TCA TCT CAG TCT CTA ATC CAG ACC ATC AAG AAT GAC ATA GTT GGA CTT AAA

+1  Arg-Asn-Leu-Asp-Ser-Arg-Met-Asn-Gln-Leu-Glu-Thr-Lys-Val-Asp-Arg-Ile-Leu-Ser-Ser-Gln-Ser-Leu-Ile-Gln-Thr-Ile-Lys-Asn-Asp-Ile-Val-Gly-Leu-Lys-(261)

          .860            .880            .900            .920            .940
GCA GGG ATG GCT ACT TTA GAA GGA ATG ATT ACA ACT GTG AAA ATC ATG GAC CCG GGA GTT CCC AGT AAT GTT ACT GTG GAA GAT GTA CGC AAG ACA CTA AGT AAC

+1  Ala-Gly-Met-Ala-Thr-Leu-Glu-Gly-Met-Ile-Thr-Thr-Val-Lys-Ile-Met-Asp-Pro-Gly-Val-Pro-Ser-Asn-Val-Thr-Val-Glu-Asp-Val-Arg-Lys-Thr-Leu-Ser-Asn-(296)

          .960            .980            .1000           .1020           .1040
CAT GCT GTT GTT GTG CCA GAA TCA TTC AAT GAT AGT TTC TTG ACT CAA TCT GAA GAT GTA ATT TCA CTT GAT GAG TTG GCT CGA CCA ACT GCA ACA AGT GTT AAG

+1  His-Ala-Val-Val-Val-Pro-Glu-Ser-Phe-Asn-Asp-Ser-Phe-Leu-Thr-Gln-Ser-Glu-Asp-Val-Ile-Ser-Leu-Asp-Glu-Leu-Ala-Arg-Pro-Thr-Ala-Thr-Ser-Val-Lys-(331)

          .1060           .1080           .1100           .1120           .1140
AAG ATT GTC AGG AAG GTT CCT CCT CAG AAG GAT CTG ACT GGA TTG AAG ATT ACA CTA GAG CAA TTG GCA AAG GAT TGC ATC AGC AAA CCG AAG ATG AGG GAA GAG

+1  Lys-Ile-Val-Arg-Lys-Val-Pro-Pro-Gln-Lys-Asp-Leu-Thr-Gly-Leu-Lys-Ile-Thr-Leu-Glu-Gln-Leu-Ala-Lys-Asp-Cys-Ile-Ser-Lys-Pro-Lys-Met-Arg-Glu-Glu-(366)

          .1160           .1180           .1200           .1220           .1240           .1260
TAT CTC CTC AAA ATC AAC CAG GCT TCC AGT GAG GCT CAG CTA ATT GAC CTC AAG AAA GCA ATC ATC CGC AGT GCA ATT TGA TCA AGA AAC ACC CAA TTA CAC TAC

+1  Tyr-Leu-Leu-Lys-Ile-Asn-Gln-Ala-Ser-Ser-Glu-Ala-Gln-Leu-Ile-Asp-Leu-Lys-Lys-Ala-Ile-Ile-Arg-Ser-Ala-Ile. (392) C-Terminus P

          .1280           .1298
ACT GTT ATG ACA CTG TAC TAA CCC TGA GGG TTT TAG A
```

Figure 1. Nucleotide Sequence of Clone P203-1 in the mRNA Sense and the Predicted Amino Acid Sequence of the Two Open Reading Frames

Nucleotide 1 is the 5' terminal nucleotide of the P and V mRNAs. After nucleotide 1298 there is a stretch of A residues in clone P203-1 (not shown) that is thought to represent part of the poly(A) tail on the mRNA. The amino acid numbering of the +1 reading frame is adjusted to conform with the residues predicted to exist in the P protein (see Figure 7B). The sequence has been deposited in the EMBL/GenBank data base (accession no. J03142).

and is part of the transcriptase complex (Buetti and Choppin, 1977), while protein V is found in infected cells and is of unknown function (Peluso et al., 1977).

The mechanism by which both the SV5 P and V proteins are encoded by a single gene has been investigated and we report here that P and V are amino coterminal proteins with different C-termini and are encoded by two separate mRNAs that differ by two nontemplated nucleotides.

Results

Nucleotide Sequence of SV5 P- and V-Specific cDNAs

To investigate the coding strategy used to express both the P and V proteins from a single gene on the SV5 virion RNA, the complete nucleotide sequence of three independently derived cDNA clones (P203-1, P10, and P127) was obtained using both dideoxynucleotide chain-terminating
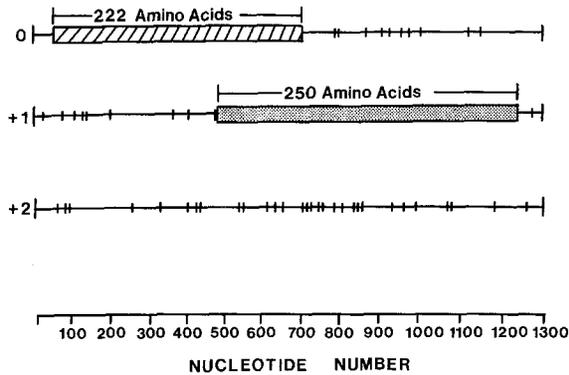
Figure 2. Schematic Representation of the Three Reading Frames in Clone P203-1

The overlapping open reading frames in the 0 and +1 reading frames are indicated by the hatched and stippled boxes respectively. The 5' and 3' ends of the clone are defined by the large vertical bars, and the termination codons in all three reading frames are indicated by the small vertical bars.

and chemical sequencing methods. The nucleotide sequence of clone P203-1 is presented in Figure 1 in the mRNA sense. It is 1298 nucleotides in length and contains an untranslated region of 60 nucleotides preceding the first AUG codon at nucleotides 61–63. Primer extension nucleotide sequencing on mRNA isolated from SV5-infected cells indicated that nucleotide 1 is the 5' terminal nucleotide of the mRNA (data not shown). At the 3' end after nucleotide 1298 in the different clones, there is a stretch of A residues of variable length. The open reading frame following the first AUG codon (reading frame 0) is capable of encoding a protein of 222 amino acids. In the

+1 reading frame there is an overlapping open reading frame of 250 amino acids, as illustrated schematically in Figure 2. Although either reading frame could encode protein V ($M_r \approx 24,000$), neither of the open reading frames is apparently large enough to encode P, a protein of $M_r \approx 44,000$ (Paterson et al., 1984), assuming the electrophoretic mobility of P is not aberrant. Examination of the predicted amino acid sequences indicates a region from residues 190–218 (29 amino acids) in the 0 reading frame that contains 7 cysteine residues, whereas the +1 reading frame only contains 1 cysteine at residue 357. Conversely, the +1 reading frame contains 6 methionine residues, whereas the 0 reading frame contains only one methionine, in addition to the initiation methionine.

P and V Monoclonal Antibody Specificity and Infected Cell Selective Labeling Patterns

To facilitate the elucidation of the coding strategy for P and V, we used monoclonal antibodies specific for the SV5 P and V proteins, which were generously made available by Dr. Rick Randall (Randall et al., 1987). The P and V monoclonal antibodies have been assigned to three groups, members of which recognize three nonoverlapping antigenic sites (R. Randall, personal communication), here designated groups I, II, and III. The proteins immunoprecipitated from SV5 infected cell lysates by a representative member of each group are shown in Figure 3 (left section). It can be seen that while P is immunoprecipitated by antibodies from all three groups, protein V is only recognized by the monoclonal antibody from group I. Recognition of both P and V by group I monoclonal antibodies indicates that they have amino acid sequences in common. The identity of the protein indicated by a star, migrating in a po-
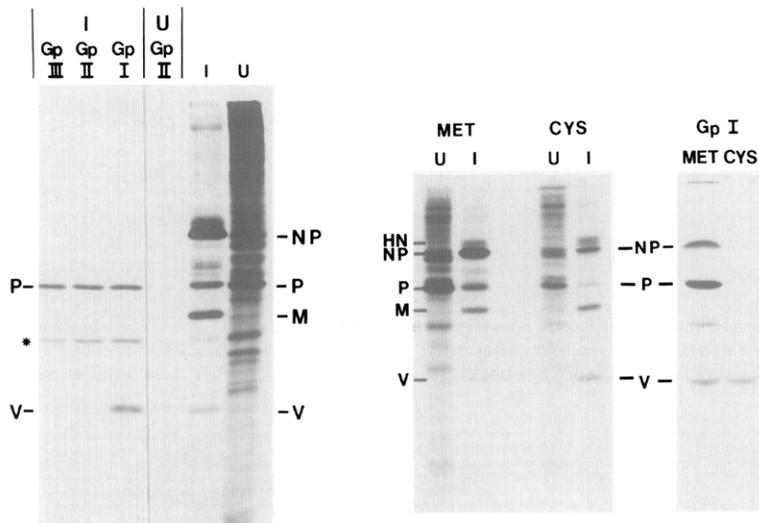
Figure 3. Characterization of P and V Specific Monoclonal Antibodies and Selective Radiolabeling of SV5-Infected Cell Proteins Using [35S]Methionine or [35S]Cysteine

Left section: SV5-infected CV1 cells were labeled at 18 hr postinfection for 1 hr with 100 μCi/ml Tran[35S]label, the cells lysed, and the labeled proteins immunoprecipitated with the P specific monoclonal antibodies and analyzed by SDS–PAGE on 15% gels as described in Experimental Procedures. I = SV5-infected cell lysate, U = uninfected cell lysate; GpI, GpII, and GpIII indicates immunoprecipitation with a Group I, II, or III P monoclonal antibody respectively.

Middle section: SV5-infected CV1 cells were labeled at 18 hr postinfection for 1 hr with either 100 μCi/ml [35S]methionine or 150 μCi/ml [35S]cysteine and the labeled proteins analyzed as above. U = uninfected cell lysate; I = SV5-infected cell lysate; MET = proteins labeled with [35S]methionine; CYS = proteins labeled with [35S]cysteine.

Right section: Immunoprecipitation of SV5-infected cell lysates using the group I monoclonal antibody. It can be seen in this immunoprecipitation that some NP coprecipitates with P and V using the group I antibody, probably because L, NP, P, and V exist in a complex (Randall et al., 1987; our unpublished data).

HN = hemagglutinin-neuraminidase protein, $M_r \approx 70,000$; NP = nucleoprotein, $M_r \approx 61,000$; P = phosphoprotein, $M_r \approx 44,000$; M = matrix protein, $M_r \approx 38,000$; V = protein V, $M_r \approx 24,000$; * = polypeptide of unknown origin.
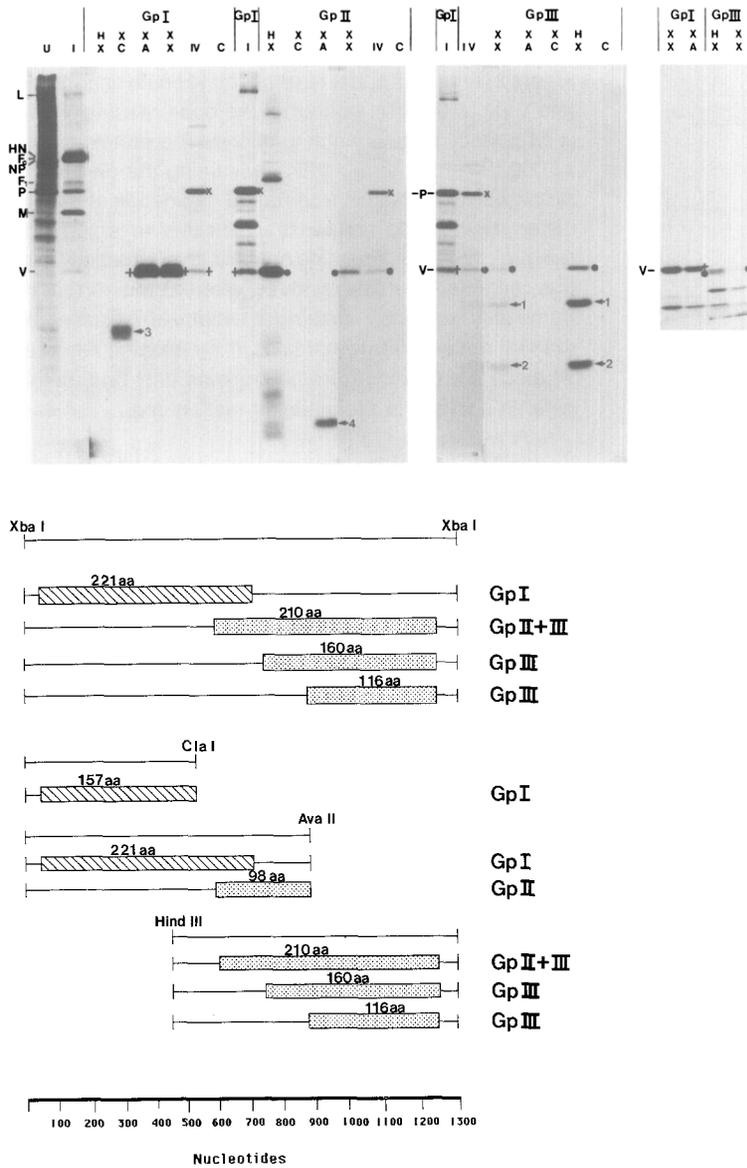
Figure 4. Antigenic Region Mapping of the P and V Monoclonal Antibodies on the P and V Gene Using T_7 RNA Polymerase Runoff Transcripts, In Vitro Translation, and Immunoprecipitation

The full-length P203-1 DNA insert cloned using XbaI linkers (XX) and a HindIII to XbaI fragment (HX) containing the 3' two-thirds of the P203-1 DNA were placed under the control of the T_7 promoter in the plasmid pGEM-2. The template DNA was linearized downstream of the T_7 promoter and the insert DNA by endonuclease digestion using EcoRI (XX and HX) or for truncated transcripts, by digestion of the full-length insert with AvaII (XA) or ClaI (XC), which cut at sites within the protein coding region. The runoff transcripts were translated in vitro using a rabbit reticulocyte lysate and the [^{35}S]methionine labeled proteins immunoprecipitated with the P specific monoclonal antibodies using the method described (Erickson and Blobel, 1979). (A) The immunoprecipitated in vitro translation products were analyzed by SDS–PAGE on 15% gels except for the 4 lanes at the far right, where a 10% polyacrylamide gel was used (see text). U = uninfected cell lysate labeled with Tran[^{35}S]label; I = infected cell lysate labeled with Tran[^{35}S]label as a marker lane; IV = proteins translated from poly(A)-containing mRNA isolated from SV5 infected cells; C = in vitro translation carried out in the absence of added mRNA; Gp I, Gp II, and Gp III indicate the monoclonal antibody used to immunoprecipitate the proteins observed in the region of the gel defined by the vertical lines either side of the antibody. Lanes XX = RNA transcribed from full-length cloned DNA; lanes HX = RNA transcribed from HindIII to XbaI 3' two-thirds fragment; lanes XA = RNA transcribed from XbaI to AvaII DNA fragment; lanes XC = RNA transcribed from XbaI to ClaI DNA fragment. X = protein P, + = protein V, ● = protein consistent in size with internal initiation at Met. 183 (210aa product in Figure 4B), 1→ = protein consistent in size with internal initiation at Met. 233 (116aa product in Figure 4B), 2→ = protein consistent in size with internal initiation at Met. 277, 3→ and 4→ = truncated products with protein synthesis initiated at Met. 1 (157aa and 98aa products respectively in Figure 4B).

(B) Schematic representation of the data accumulated in Figure 4A showing the protein products derived from the 0 and +1 reading frames, their relative number of amino acids assuming that all protein products from the +1 reading frame are initiated from internal methionine residues, the restriction endonuclease sites used in the generation of the RNAs, and a summary of the protein product reactivity with the Gp I, II, and III monoclonal antibodies.

sition between that of P and V, has not been investigated. To examine the relative ability by which the P and V proteins could be selectively radiolabeled, SV5-infected cells were labeled with either [^{35}S]methionine or [^{35}S]cysteine (Figure 3, middle section). In addition, cell lysates were immunoprecipitated with the group I antibody (Figure 3, right section). As shown in Figure 3, protein V was easily detected when labeled with either methionine or cysteine, whereas the P protein, although readily labeled with methionine, was poorly labeled with cysteine. These observations suggest that protein V possesses the cysteine-rich region encoded by the 0 reading frame, while the P protein apparently does not.

Mapping Coding Regions for P and V

To define the regions of the nucleotide sequence encoding the P and V proteins, we used the approach of making synthetic mRNA transcripts, translating the RNAs in vitro, and immunoprecipitating the products. The complete coding regions of clone P203-1 (nucleotides 44–1285) and a fragment containing the 3' two-thirds of the gene were subcloned into the transcription vector pGEM-2 such that they were under the control of the T_7 RNA polymerase promoter. A series of mRNA-sense runoff transcripts were prepared as described in Experimental Procedures, translated in vitro using a rabbit reticulocyte lysate, and immunoprecipitated using the group I, II, and III monoclonal an-

tibodies. The results obtained from such an assay are shown in Figure 4A and summarized in schematic form in Figure 4B.

Interestingly, although the P protein could be translated in vitro using poly(A)-containing mRNAs isolated from SV5 infected cells (Figure 4A, indicated as x in lanes IV), we were unable to detect the synthesis of the P protein when in vitro runoff transcripts were used to program the cell-free translation system. However, protein V was translated from both synthetic RNAs and mRNA was isolated from infected cells (Figure 4A, indicated as +). Originally we thought it likely that frameshifting might be involved in the synthesis of P. However, because there is no detectable synthesis of the P protein when T_7 runoff transcripts are used to program the rabbit reticulocyte lysate, whereas P is translated efficiently in vitro when poly(A) containing mRNA from infected cells is used, it would seem unlikely that ribosomal frameshifting is involved in the generation of the P protein. In addition to protein V, other in vitro synthesized proteins were observed, particularly during translation of the 3′ two-thirds of the gene. The apparent size of the additional proteins is consistent with initiation of protein synthesis occurring at internal AUG codons in the +1 open reading frame: methionines 183 (Figure 4A, closed circle), 233 (Figure 4A, 1→), and 277 (Figure 4A, 2→). From the sizes of the different protein products derived from the various runoff transcripts, it was possible to map unambiguously the regions on the open reading frames that were recognized by the monoclonal antibodies. In this way, monoclonal antibodies from group I were found to recognize the N-terminal region of the 0 open reading frame, group II antibodies to recognize a region from the N terminus of the +1 open reading frame, and group III antibodies to recognize a C-terminal region of the +1 open reading frame.

In this analysis it should be noted that the largest in vitro translation product recognized by group II and III antibodies (Figure 4A, indicated by a closed circle) was very similar in size to protein V (210 amino acids versus 221 amino acids) and had an almost identical electrophoretic mobility on SDS–PAGE. However, these polypeptides could be resolved when the samples were analyzed on a 10% polyacrylamide gel (Figure 4A, right four lanes).

Thus, these data indicate that because P and V are immunoprecipitated by the Gp I monoclonal antibodies, and V cannot be immunoprecipitated by the Gp II and Gp III monoclonal antibodies, P and V are amino coterminal and protein V must be the product of the 0 open reading frame. In addition, these data indicate that protein P is derived from amino acid residues encoded by a large part of the +1 reading frame.

Elimination of the Protein V Termination Codon

To obtain additional evidence that protein V is encoded by the 0 open reading frame and that the stop codon (TAA, nucleotides 727–729) that terminates translation in the 0 reading frame is not an artifact of the cDNA cloning procedure, we used an approach involving site-specific mutagenesis. If this stop codon in used to terminate translation of protein V, its elimination should prevent a normal-sized
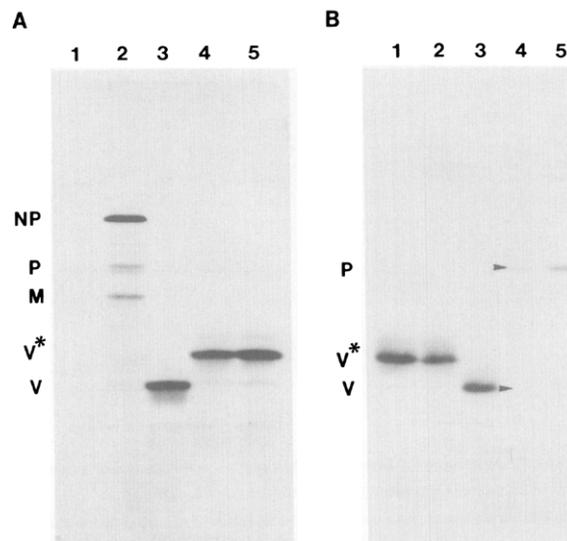
Figure 5. Elimination of the Protein V Termination Codon

The termination codon (TAA) at nucleotides 727–729 was changed by oligonucleotide-directed mutagenesis to GCG, RNA was transcribed from the mutated template DNA and translated in vitro in a rabbit reticulocyte lysate as described in the legend to Figure 4.

(A) The in vitro–synthesized proteins were analyzed by electrophoresis on a 15% SDS polyacrylamide gel. Lane 1 = no RNA control translation; Lane 2 = translation of poly(A)-containing mRNA from SV5 infected cells; Lane 3 = translation of a T_7 RNA polymerase transcript of clone P203-1; Lanes 4 and 5 = translation of T_7 RNA polymerase transcripts containing the stop codon mutation. NP = nucleoprotein; P = phosphoprotein; M = matrix protein; V* = extended mutant protein V; V = protein V.

(B) The products of in vitro translation were immunoprecipitated using the group I monoclonal antibody and analyzed as described in (A) Lanes 1 and 2 = immunoprecipitation of mutant V* protein from the translation products shown in (A), lanes 4 and 5; Lane 3 = immunoprecipitation of protein V from translation products shown in (A), lane 3; Lane 4 = immunoprecipitation of the proteins synthesized in SV5-infected cells labeled with Tran[^{35}S]label; Lane 5 = immunoprecipitation of P and V proteins from translation products shown in (A), lane 2.

protein V from being synthesized, and a larger protein of 254 amino acids should be found. Nucleotides 727–729 (TAA) in the cloned DNA were changed to the triplet GCG (encoding alanine) as described in Experimental Procedures, the DNA containing the mutation was transcribed in vitro, and the resulting synthetic RNA translated in a rabbit reticulocyte lysate. As shown in Figure 5A, lanes 4 and 5, a protein (V*) larger than V (Figure 5A, lane 3) that was recognized by the group I monoclonal antibody (Figure 5B, lanes 1 and 2) was synthesized.

Identification of a Second mRNA Species

As no evidence for frameshifting could be obtained, i.e., the inability to translate P in vitro from T_7 transcripts of P203-1 cloned DNA, the most plausible mechanism by which P and V are encoded is that a second mRNA species that is translated to yield the P protein exists. An insertion or deletion in such a mRNA would be expected to occur in the region of overlap between the two reading frames shown in Figure 2. To search for the existence of a second mRNA population, nuclease S1 protection anal-
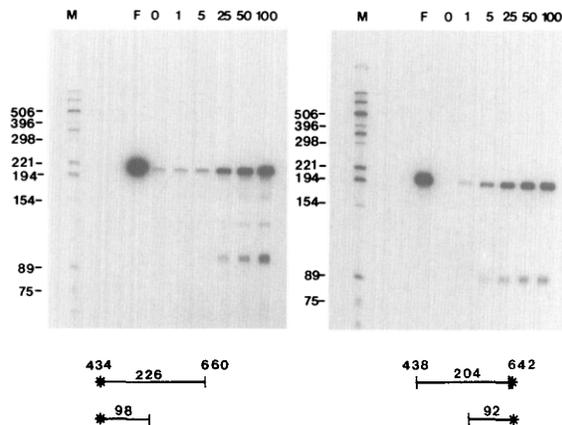
Figure 6. Nuclease S1 Analysis to Assay for the Existence of a Second mRNA Species Originating from the P and V Gene

Two DNA fragments, one uniquely 3' end-labeled at nucleotide 434 and the other uniquely 5' end-labeled at nucleotide 642, were hybridized to increasing amounts of poly(A)-containing mRNA isolated from SV5 infected cells and the hybrids digested with nuclease S1. The protected fragments were analyzed by electrophoresis on 6% acrylamide gels containing 9M urea. Lanes: M, [^{32}P]labeled size markers from HinfI-digested pFVM45 DNA (Lamb and Lai, 1982); F, control untreated probe; 0, no added mRNA; 1–100, increasing mRNA concentrations in the ratio 1:5:25:50:100. Numbers on the left of each panel are nucleotide sizes. A schematic diagram of the probe protected products, their nucleotide sizes, and the position of the uniquely labeled end, which is indicated by a star is shown beneath the autoradiograms.

ysis was performed using poly(A)-containing mRNA from SV5 infected cells and two DNA fragments from clone P203-1 that spanned the entire region of overlap, one uniquely 3' end-labeled at nucleotide 42 and the other uniquely 5' end-labeled at nucleotide 892. With each fragment, two nuclease S1 protected labeled fragments were detected, the full-length fragment used as a probe corresponding to a colinear mRNA transcript, and a smaller fragment present in 10%–20% abundance (data not shown). These data suggest that a second mRNA species exists that is derived from the P and V gene on the SV5 virion RNA and that it has a nuclease S1 sensitive site approximately between nucleotides 530 and 560. To define the location of this site more accurately, two shorter DNA fragments were used for the nuclease S1 analysis. One DNA fragment (nucleotides 434 to 660) was 3' end-labeled at nucleotide 434 and the other DNA fragment (nucleotides 438 to 642) was 5' end-labeled at nucleotide 642. In addition to protection of the probe fragments corresponding to a colinear mRNA transcript, both probes protected smaller fragments found in 10%–20% abundance (Figure 6). With both the 3' end and 5' end-labeled probes, smaller protected fragments (98 and 92 nucleotides respectively) increased in abundance with increasing concentrations of mRNA (Figure 6). The size of the protected fragments mapped the region containing the nuclease S1-sensitive region to between nucleotides 532 and 550.

Isolation of cDNAs that Contain Two Nontemplated Nucleotides and Encode Protein P

A cDNA library derived from SV5-infected CV1 cell mRNAs (Paterson et al., 1984) was screened with an oligo-

nucleotide probe to isolate cDNA clones specific for the P and V mRNAs. The nucleotide sequence of 22 P and V specific cDNA clones was obtained over the region of overlap between the 0 and +1 reading frames. In addition to cDNA clones having the same sequence as P203-1 (12 clones), a second population of cDNA clones was isolated (10 clones) that differed from P203-1 in containing two additional bases between nucleotides 548–551. The nucleotide sequence over the relevant region of a P mRNA clone and a V mRNA clone is shown in Figure 7A. It can be seen that whereas the V cDNA (P203-1) has four G residues between nucleotides 548 and 551, the P cDNA has six G residues (sections P and V, Figure 7A). Although the simplest explanation of the nuclease S1 mapping data was that the P mRNA was a noncolinear transcript of the P and V gene containing an 18 nucleotide interrupted region, our retrospective explanation for the data is that nuclease S1 recognized a two nucleotide mismatch. The 3' break point maps precisely to the 4 G region in the V cDNA clone, while the 5' site did not. In addition to the inherent inaccuracies in measuring the precise size of DNA fragments, it is noted that the region 5' to the 4 G residues at nucleotides 548–551 is AT-rich and it may have been sensitive to digestion by the nuclease S1 (Hansen et al., 1981).

The two extra G residues cause a switch from the 0 reading frame to the +1 reading frame, and the predicted amino acid sequences are shown in Figure 7B. These data indicate that the P mRNA has the capacity to encode a polypeptide of 392 amino acids initiating at the AUG codon at nucleotides 61–63 and terminating at the TGA codon at nucleotides 1237–1239.

To determine whether the genomic virion RNA from which the SV5 mRNAs are transcribed contains four or six C residues complementary to the four or six G residues found in the two mRNAs, the sequence of the virion RNA (vRNA) was obtained as described in Experimental Procedures. As shown in Figure 7 (section vRNA), only a single cDNA sequence could be detected, and it contained four G residues complementary to four C residues in the vRNA.

Expression of P In Vitro from a P cDNA Clone

To provide further evidence that the only difference between a P mRNA and a V mRNA is the presence of two nontemplated G residues, we investigated whether the P protein could be translated from a synthetic RNA derived from a P cDNA. To facilitate the genetic manipulation, an internal large restriction fragment spanning the region of interest in the P203-1 pGEM-2 vector was replaced with the comparable fragment from the P cDNA clone. In addition to using a "natural" P cDNA, we also changed the V cDNA clone P203-1 by site-specific mutagenesis to insert two additional G residues into the four G residues at nucleotides 548–551. Synthetic RNA was transcribed from both the "natural" and the "synthetic" P cDNA clones using T$_7$ RNA polymerase, and translated in vitro in rabbit reticulocyte lysates. As shown in Figure 8, both the "natural" and "synthetic" P cDNA clones yielded a P protein with an electrophoretic mobility identical to that of the P protein synthesized in infected cells and to the P protein translated in vitro from SV5-infected cell mRNA. All the P
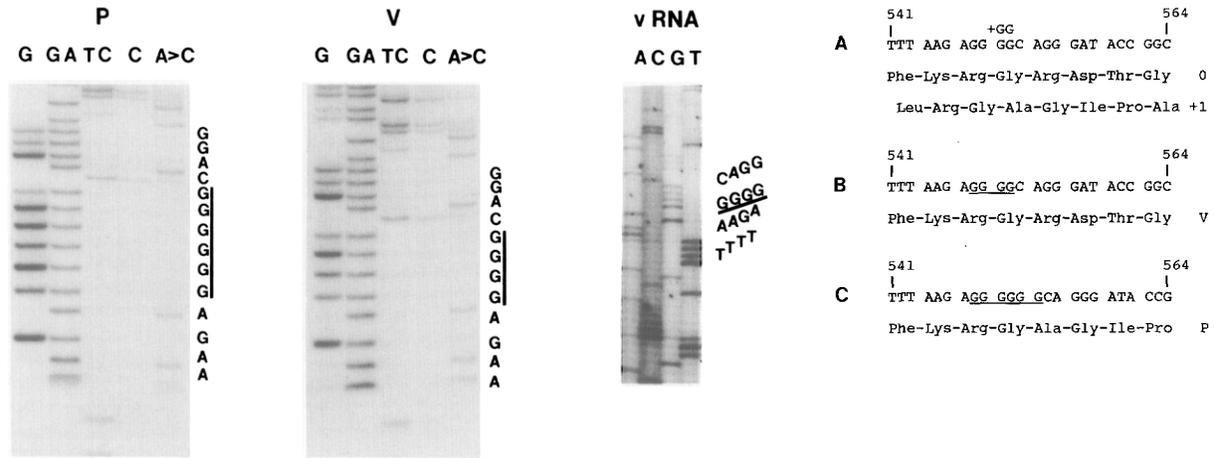
Figure 7. Sequence of a P mRNA cDNA Clone in Comparison to the Sequence of a V mRNA cDNA Clone and the Sequence of the SV5 Genomic RNA

(A) The nucleotide sequences of a P cDNA clone and a V cDNA clone in the region of nucleotides 541–564 are shown to illustrate the six G or four G residues in the P cDNA and V cDNA respectively. Sequencing was done by the chemical cleavage method (Maxam and Gilbert, 1980). The sequence of the SV5 genomic template RNA (vRNA) is shown in the message sense as determined by dideoxy primer extension sequencing using reverse transcriptase (Air, 1979).

(B) The predicted amino acid sequence of the P and V proteins in the region of the six G or four G residues.
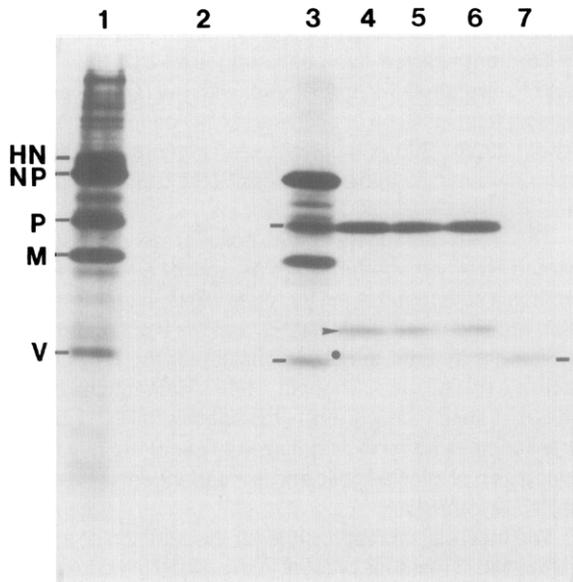
Figure 8. Expression of the P Protein from In Vitro Synthesized RNA

A P cDNA clone was reconstructed in the P203-1 pGEM2 vector by replacing an internal large PstI–DNA fragment (nucleotides 225–660) with that from a P cDNA containing the two nontemplated G residues between nucleotides 548–551. The P203-1 DNA was also changed by site-specific mutagenesis to insert two additional G residues into the four G residues at nucleotides 548–551, and the mutated DNA subcloned into the pGEM-2 vector. RNA was transcribed with T₇ RNA polymerase from both the "natural" and the "synthetic" P cDNA clones and translated in vitro using rabbit reticulocyte lysates. Lane 1 = SV5-infected CV1 cell lysate as a marker. In vitro translated RNAs were as follows: lane 2 = no RNA control; lane 3 = poly(A)-containing mRNAs from SV5-infected CV1 cells; lanes 4 and 5 = "synthetic" P RNA from site-specifically mutated template DNAs; lane 6 = "natural" P RNA; lane 7 = V RNA from clone P203-1. Dashes = proteins P and V. Arrowhead and dot indicate protein products thought to originate from initiation at internal methionine residues 141 and 183 respectively.

proteins could be immunoprecipitated with the group I, II, and III monoclonal antibodies (data not shown). The protein products found in Figure 8, lanes 4–6, indicated by an arrow and a dot, are thought to be internal initiation products from methionine residues 141 and 183 respectively. Protein V (Figure 8, lanes 3 and 7) is of a slightly different electrophoretic mobility from the smaller internal initiation product, and only protein V and not the internal initiation products are precipitated by the group I monoclonal antibodies (data not shown).

RNA Secondary Structure Predictions

The finding of two extra G residues at a precise location in the P mRNA suggests that a signal would be needed to specify their addition, such as a region of strong secondary structure in the vRNA or mRNA. With the aid of the computer program FOLD (Intelligenetics Inc., Palo Alto, CA), the most stable secondary structure that can be predicted for nucleotides 520–620 of the P and V gene is one with an energy of $\Delta G = -53.7$ kcal/mol and has the four templated C residues (nucleotides 548–551) immediately after a base-paired stem region (Figure 9).

The Klenow fragment of E. coli DNA polymerase often yields artifactual sequencing bands at a run of several G residues when directly sequencing double-stranded DNA using the dideoxy chain-terminating method. As shown in Figure 9 the sequence of nucleotides 548–551 in the V clone (4 G residues) is easier to interpret than nucleotides 548–553 in the P clone (6 G residues). These artifactual bands can be eliminated when the sequencing is performed with a modified form of T₇ DNA polymerase (Sequenase™) in conjunction with dITP instead of dGTP, unless there is a strong secondary structure in the template strand and then the artifacts are exacerbated (Tabor, 1987). When this was done for the P clone DNA (Figure

```
           520              620
           vRNA sense
      3'-UAGCGU   GUUCCA-5'
              U·A
              G-C
                 U
              G-C
              U·A
                 U
              A·U
                 C
             A  GC
                 C
              GC
              GC
              GC
              A·U
              A·U
              GC
              CGC
              U·A
              A·U
            A-A-A·U
            U-U-CG
                 U·A
                 CGG
      CCCUG CCC  UACUCCCU
     U                    U
     A                     C
      U GGCCGCCCAAGGUAU
```

TRANSCRIPTION ↓
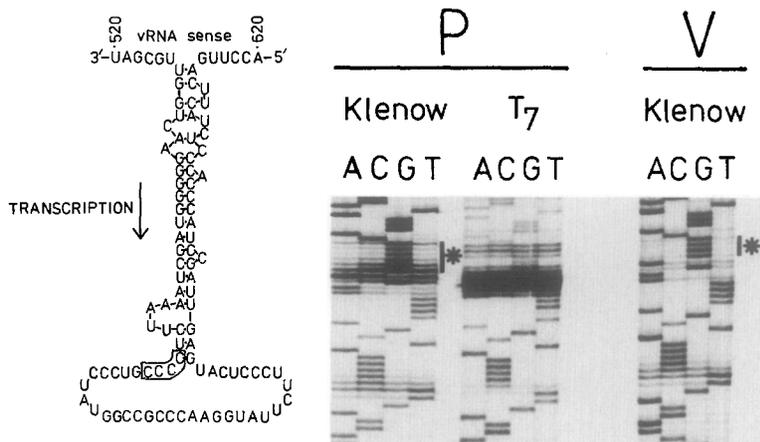
Figure 9. RNA Secondary Structure Predictions

Left: The SV5 virion RNA sequence from nucleotides 520–620 of the P/V gene was examined for regions of strong secondary structure with the aid of the computer program FOLD (Intelligenetics Inc., Palo Alto, CA). The stem–loop structure shown has an energy of $\Delta G = -53.7$ kcal/mol. The four C residues at nucleotides 548–551 are boxed. The arrow denotes the direction of mRNA transcription.

Right: Nucleotide sequences obtained by the dideoxynucleotide chain-terminating method using the Klenow fragment of E. coli DNA polymerase or a modified form of T_7 DNA polymerase (Sequenase™) on a P clone cDNA template (Klenow and T_7) or a V clone cDNA template (Klenow). In the Sequenase reactions (T_7) dITP was used in place of the usual dGTP. The region of the four or six G residues between nucleotides 548–551 is indicated by a star.

9) or V clone DNA (data not shown), the T_7 DNA polymerase nearly stopped its processive synthesis at nucleotides 543–550, which suggests that there is a native secondary structure in this region.

Discussion

We have obtained the nucleotide sequence of the paramyxovirus SV5 P and V gene and have determined the strategy by which both proteins are expressed by a single gene. The P and V proteins are translated from two independent mRNAs that are synthesized in SV5 infected cells and are found to differ by the presence in the P mRNA of two additional nucleotides. A comparison of the nucleotide sequences of the P and V cDNAs and the SV5 genomic vRNA showed that the two additional G residues present in the P mRNA are not templated by the SV5 virion RNA (Figure 7). It could be argued that the vRNA sequencing might not detect a minor vRNA species of less than 5% abundance. However, there is no biological evidence for the involvement of more than one virus genome in the SV5 infectious cycle. Using a combination of in vitro translation of T_7 runoff transcripts, immunoprecipitation of the in vitro synthesized proteins using monoclonal antibodies, oligonucleotide-directed mutagenesis, and metabolic labeling of SV5 infected cell proteins using specific amino acids ([^{35}S]methionine or [^{35}S]cysteine), we have shown that P and V are amino coterminal proteins that have different C-termini. The results presented here confirm earlier observations that the P and V proteins of SV5 have tryptic peptides in common (Paterson et al., 1984). Thus SV5 differs from many paramyxoviruses and morbilliviruses that use functionally bicistronic mRNAs to synthesize the P protein, and a second protein known as C from overlapping reading frames (Giorgi et al., 1983; Galinski, et al., 1986; Bellini et al., 1985; Barrett et al., 1985). Early peptide mapping data obtained for the P and "C-like" proteins of two other paramyxoviruses, Newcastle

disease virus (NDV) and mumps virus, suggested that both proteins are encoded by the same reading frame (Collins et al., 1982; Herrler and Compans, 1982). Recently the NDV and mumps virus P genes have been sequenced and found to contain one open reading frame (Sato et al., 1987; McGinnes et al., 1988; Takeuchi et al., 1988) from which it has been suggested that both the P and "C-like" proteins are derived, with the "C-like" protein arising from initiation at an internal AUG codon (McGinnes et al., 1988). SV5 is therefore seemingly unique among paramyxoviruses in having two mRNAs transcribed from the P gene.

The RNA-dependent RNA polymerase of negative strand RNA viruses functions as part of a transcriptase complex composed of the template vRNA in tight association with the nucleoprotein (NP), and the P and L proteins, which are thought to be responsible for the polymerase activity (Buetti and Choppin, 1977; Hamaguchi et al., 1983). Transcription of the virus-specific mRNAs by the transcriptase complex is believed to occur entirely in the cytoplasm of infected cells and is independent of host–cell mRNA synthesis.

The mechanism responsible for the addition of the untemplated G residues present in the P mRNA is unknown, nor is it known whether it is a cotranscriptional or posttranscriptional process. However, the virus-encoded RNA polymerase of negative strand RNA viruses is also responsible for the polyadenylation of virus-specific mRNAs, a process that is thought to occur by a "slippage" or "stuttering" mechanism involving the reiterative copying by the polymerase of a stretch of U residues located at the end of each gene. As the nontemplated G residues are added to the P transcript at a position where the template vRNA has a run of four C residues, it is possible that the SV5 polymerase "stutters" while copying this region of the genome and thus adds the nontemplated nucleotides. It is interesting that immediately upstream of the four C residues on the SV5 genomic RNA is the sequence

SV5	562-726	G G F H R R E Y S I G W V G D E V K V T E W C N P S C S P I T A A A R R
MUMPS	663-833	G G - H R R E Y S L S Y V Q G E V R V F E W C N P I C S P I T A A A R F
NDV	605-799	P G - H R R E H T I S W T M G G V T T I S W C N P S C S P I R A E P R Q
SENDAI	1048-1254	K G - H R R E H I I Y E R D G Y I V D E S W C N P V C S R I R I I P R R
PI-3	1132-1332	E G - Y R R E Q S I Y K E G N Y S I A E S W C N S I Y T K T R F I S R Q
MEASLES	749-958	K G - H R A Q I S L I W N G D R V F I D R W C N P M C S K V T L G T I R

SV5	F E C T C H Q C P V T C S E C E R D T *
MUMPS	H S C K C G N C P A K C D Q C E R D Y G P P *
NDV	Y S C T C G S C P A T C R L C A G D D L Y D G G D I T E G K *
SENDAI	E L C V C K T C P K V C K L C R D D I Q C M R P D P F C R E T F R S *
PI-3	T S C M C S K C T K Q C R Y C I K D R L P S R I S H R G F N G Q *
MEASLES	A R C T C G E C P R V C E Q C R T D T G V D T R I W Y H N L P E I P E *

Figure 10. Amino Acid Sequence Homology Between the Cysteine-Rich Region of Protein V of SV5 and Amino Acid Sequences Predicted from Open Reading Frames of Other Paramyxovirus P Genes

The published nucleotide sequences of the P genes of several paramyxoviruses and the morbilivirus, measles virus, were translated in all three reading frames. In each case of a reading frame overlapping that for the P protein a cysteine-rich region was identified and is listed in the single letter amino acid code. Only the region of significant conservation of sequence is shown with its corresponding nucleotide number; the N-terminal region of the open reading frame is omitted. The star at the end of the amino acid sequence represents a translation termination codon. The boxes identify positions where three or more amino acids have been conserved in all six viruses. A dash indicates that a gap was placed in the alignment, and the star above the sequences shows the seven conserved cysteine residues. Sources for the P gene nucleotide sequences are as follows: SV5, this publication; mumps virus, Takeuchi et al., 1988; NDV, Sato et al., 1987; Sendai virus, Shioda et al., 1983 and Giorgi et al., 1983; parainfluenza virus 3 (PI-3), Galinski et al., 1986, and Luk et al., 1986; measles virus, Bellini et al., 1985.

3'-AAAAUUCU-5' (Figure 9), which resembles the putative polyadenylation signal found at the end of SV5 genes and in fact is identical to the sequence at the end of the SV5 HN gene (Hiebert et al., 1985), making this an attractive model for the mechanism by which the nontemplated Gs are added. However, it cannot be ruled out that the nontemplated G residues in the P mRNA are added as a consequence of some form of RNA-editing analogous to that found in mitochondrial transcripts in trypanosomes (Benne et al., 1986; Feagin et al., 1987, 1988; Shaw et al., 1988) or the mammalian apolipoprotein-B mRNA (Powell et al., 1987; Chen et al., 1987). While screening the SV5 cDNA library for a P cDNA, 22 clones were sequenced across the region described above and only clones with either four or six G residues were found. This would suggest that whatever the mechanism involved in the addition of the nontemplated G residues, it is extremely specific. With the aid of the computer algorithm FOLD, a region of secondary structure was predicted for this part of the template RNA and it is therefore possible that this could play a role in either the mechanism itself or its regulation (Figure 9).

An examination of the predicted amino acid sequences of the P and V proteins reveals several interesting features. As mentioned above, the P and V proteins are amino coterminal and have their first 164 residues in common (Figure 1). An unusual feature of the shared region is the large number of proline residues; 17 prolines in 164 amino acids (Figure 1). However, the most striking characteristic observed in either protein is the C-terminal portion of protein V, which consists of a cysteine-rich region bearing a remarkable resemblance to cysteine-rich regions found in the adenovirus E1A protein (for review see Moran and Matthews, 1987), the yeast transcription factor GAL4 (Johnston and Dover, 1987), and proteins belonging to the steroid hormone receptor superfamily (for review see Evans, 1988). In these proteins and others possessing a similar domain it is thought that the binding of metal ions by the cysteine-rich region plays an important role in either the binding of nucleic acid by the protein, mediat-

ing protein–protein interactions, or stabilizing oligomeric forms of a protein, as in the Tat protein of human immunodeficiency virus (Frankel et al., 1988). Because of the significance of the cysteine-rich regions in other proteins, it was of interest to determine whether the sequence identified here in protein V had been conserved among other paramyxovirus P genes. Consequently, we compared the cysteine-rich region from protein V with the protein sequences predicted in all three reading frames from the nucleotide sequence of the P genes from mumps virus, NDV, Sendai virus, parainfluenza virus 3, and measles virus (Takeuchi et al., 1988; Sato et al., 1987; McGinnes et al., 1988; Shioda et al., 1983; Giorgio et al., 1983; Galinski et al., 1986; Luk et al., 1986; Bellini et al., 1985). As shown in Figure 10, a highly conserved cysteine-rich region was identified in an open reading frame in all the different paramyxovirus P gene sequences examined. Interestingly, the cysteine-rich region is more conserved between the different paramyxoviruses than is the amino acid sequence of the P protein encoded by the same nucleotides but translated in another reading frame (data not shown). As the P protein is part of the paramyxovirus transcriptase complex, the conservation of the cysteine-rich region must have important biological significance. It will be important to determine whether a protein containing this cysteine-rich region is synthesized in cells infected with other paramyxoviruses in addition to the already identified P or P and C proteins derived from the "P" gene.

The function of protein V has yet to be elucidated. However, as V is found associated with purified SV5 virions (our unpublished data) and as group I antibodies precipitate L, NP, P, and V in a complex (Randall et al., 1987; our unpublished data), it remains a possibility that protein V may play a role in transcription and/or replication of the virus genome in infected cells.

Experimental Procedures

Cells
Monolayer cultures of a variant of the MDBK line of bovine kidney cells

and the TC7 clone of CV-1 cells were grown in Dulbecco's modified Eagle's medium (DmEm) supplemented with 10% fetal calf serum.

Virus Infections and Labeling of Cells

Stock virus was grown in MDBK cells infected with the W3 strain of SV5 (Choppin, 1964) as described previously (Peluso et al., 1977). For all biochemical experiments, CV-1 cells were used and infected as described previously (Paterson et al., 1984), except that for metabolic labeling of infected cell proteins, monolayers were incubated in methionine- and cysteine-free DmEm and proteins labeled using either Tran[^{35}S]label (ICN Radiochemicals, Irvine, CA), [^{35}S]cysteine or [^{35}S]methionine (Amersham Corp., Arlington Heights, IL).

Isolation of Poly(A)-Containing mRNAs

Messenger RNAs were isolated as described previously (Paterson et al., 1984).

Molecular Cloning of SV5 mRNAs and Nucleotide Sequencing

cDNA synthesis, isolation of SV5 specific clones, and the identification of cDNA encoding the various viral gene products has been described (Paterson et al., 1984). Three clones, P10, P27, and P203-1 were sequenced over their entire length both by the chemical cleavage method (Maxam and Gilbert, 1980) and after subcloning into the PstI site of the replicative form of bacteriophage M13mp19, by the dideoxy chain-termination method (Sanger et al., 1977). Dideoxy primer extension sequencing on purified SV5 genomic RNA and poly(A)-containing mRNA was performed using avian myeloblastosis virus reverse transcriptase (Molecular Genetic Resources, Tampa, FL) and P gene specific primers as described previously (Air, 1979). Direct sequencing of double-stranded plasmid DNA was carried out by the dideoxy chain-termination method using the Klenow fragment of E. coli DNA polymerase (Bethesda Research Laboratories, Gaithersberg, MD) as described by Sanger et al. (1977) or a modified form of T$_7$ polymerase (Sequenase^{TM}, United States Biochemical Corp., Cleveland, OH) according to the manufacturer's instructions. Restriction endonucleases, bacterial alkaline phosphatase, and T$_4$ DNA ligase were obtained from Bethesda Research Laboratories, and T$_4$ polynucleotide kinase from Pharmacia Fine Chemicals (Piscataway, NJ). Oligonucleotides were synthesized by the Northwestern University Biotechnology Facility on an Applied Biosystems (Foster City, CA) model 380B DNA synthesizer and were purified as described (Paterson and Lamb, 1987).

Site-Specific Mutagenesis and Plasmid Construction

The P203-1 cDNA was excised from pBR322 by HhaI and MstII digestion, thereby eliminating the G/C tails introduced during cDNA cloning; XbaI linkers were added and the cDNA subcloned into the XbaI site of pGEM-2 (Promega Biotec, Madison, WI). Deletion of the 5' end of the gene was performed by digesting pGEM-2 containing the P203-1 cDNA with XbaI and HindIII, isolating the 3' portion of the gene, addition of XbaI linkers, and subcloning back into pGEM-2. To construct both the protein V stop codon elimination mutant and the frameshift mutant, P203-1 cDNA was subcloned into the XbaI site of the replicative form of bacteriophage M13mp19. Oligonucleotide-directed mutagenesis was carried out according to the procedure of Zoller and Smith (1982) using mutagenic oligonucleotides consisting of 12 nucleotides either side of the site of the mutation. DNA containing the desired mutation was subcloned into pGEM-2 and the mutation verified by direct plasmid DNA sequencing using the dideoxy chain termination method (Sanger et al., 1977) and a P specific oligonucleotide primer.

In Vitro RNA Synthesis

For transcription of the entire coding region, plasmid DNAs were linearized downstream of the T$_7$ promoter and the P or V insert, using EcoRI. For the synthesis of truncated forms of the mRNA, the DNA template was linearized using either AvaII or ClaI, which recognize sites within the coding region of the cDNA. In vitro synthesis of mRNA was carried out as described previously (Hull et al., 1988) and 1 μg of RNA was used to program a rabbit reticulocyte lysate as described below. T$_7$ DNA-dependent RNA polymerase was obtained from Bethesda Research Laboratories, RNasin^{TM} and RQ DNase^{TM} from Promega Biotec, and ^{7m}G(5')ppp(5')G (sodium salt) was from Pharmacia Fine Chemicals.

In Vitro Translation of mRNAs

mRNAs were translated in vitro using a micrococcal nuclease-treated rabbit reticulocyte lysate (Promega Biotec) according to the manufacturer's instructions. The in vitro-synthesized products were labeled using [^{35}S]methionine. One-fifth volume of each translation reaction was immunoprecipitated as described below.

Immunoprecipitation and SDS-PAGE

Immunoprecipitation was performed as previously described (Lamb et al., 1978; Erickson and Blobel, 1979) using monoclonal antibodies to the P and V proteins kindly provided by Dr. Rick Randall (Randall et al., 1987). Samples were prepared for electrophoresis and analyzed by SDS-PAGE on 15% polyacrylamide gels as previously described (Lamb et al., 1978).

Nuclease S1 Analysis

Poly(A)-containing mRNAs from SV5 infected CV-1 cells were isolated as described (Paterson et al., 1984). To determine whether more than one mRNA is transcribed from the P gene nuclease, S1 analysis was performed as previously described (Lamb and Lai, 1982). The labeled DNA fragments used as probes were: a HhaI–AvaII fragment and a BamHI–PstI DNA fragment (nucleotides 42–889 and 434–660, respectively) 3' uniquely labeled at nucleotides 42 and 434, and a HhaI–AvaII fragment and a BamHI–HphI fragment (nucleotides 42–892 and 438–642, respectively) 5' uniquely labeled at nucleotides 892 and 642. Nuclease S1 was obtained from Boehringer Mannheim Biochemicals, Indianapolis, IN.

References

Air, G. M. (1979). Nucleotide sequence coding for the "signal peptide" and N terminus of the hemagglutinin from an Asian (H2N2) strain of influenza virus. Virology 97, 468–472.

Atkins, J. F., Steitz, J. A., Anderson, C. W., and Model, P. (1979). Binding of mammalian ribosomes to MS2 phage RNA reveals an overlapping gene encoding a lysis function. Cell 18, 247–256.

Barrett, T., Shrimpton, S. B., and Russell, S. E. H. (1985). Nucleotide sequence of the entire protein coding region of canine distemper virus polymerase-associated (P) protein mRNA. Virus Res. 3, 367–372.

Bellini, W. J., Englund, G., Rozenblatt, S., Arnheiter, H., and Richardson, C. D. (1985). Measles virus P gene codes for two proteins. J. Virol. 53, 908–919.

Beene, R., Van Den Burg, J., Brackenhoff, J. P. J., Sloof, P., Van Boom, J. H., and Tromp, M. C. (1986). Major transcript of the frameshift coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46, 819–826.

Bertholet, C., Van Meir, E., ten Heggeler-Bordier, B., and Wittek, R. (1987). Vaccinia virus produces late mRNAs by discontinuous synthesis. Cell 50, 153–162.

Bos, J. L., Polder, L. J., Bernards, R., Schrier, P. I., van den Elsen, P. J., van der Eb, A. J., and van Ormondt, H. (1981). The 2.2 kb E1b mRNA of human Ad12 and Ad5 codes for two tumor antigens starting at different AUG triplets. Cell 27, 121–131.

Boursnell, M. E. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M., and Binns, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J. Gen. Virol. 68, 57–77.

Breitbart, R. E., Andreadis, A., and Nadal-Ginard, B. (1987). Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annu. Rev. Biochem. 56, 467–495.

Briedis, D. J., and Lamb, R. A. (1982). Influenza B virus genome: sequences and structural organization of RNA segment 8 and the mRNAs coding for the NS₁ and NS₂ proteins. J. Virol. 42, 186–193.

Brierley, I., Boursnell, M. E. G., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. K., and Inglis, S. G. (1987). An efficient ribosome frameshifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J. 6, 3779–3785.

Buetti, E., and Choppin, P. W. (1977). The transcriptase complex of the paramyxovirus SV5. Virology 82, 493–508.

Chen, S.-H., Habib, G., Yang, C.-Y., Gu, Z.-W., Lee, B. R., Weng, S. A., Silberman, S. R., Cai, S.-J., Deslypere, J. P., Rosseneu, M., Gotto, Jr., A. M., Li, W.-H., and Chan, L. (1987). Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. Science 238, 363–366.

Choppin, P. W. (1964). Multiplication of a myxovirus (SV5) with minimal cytopathic effects and without interference. Virology 23, 224–233.

Collins, P. L., Wertz, G. W., Ball, L. A., and Hightower, L. E. (1982). Coding assignments of the five smaller mRNAs of Newcastle disease virus. J. Virol. 43, 1024–1031.

Curran, J. A., Richardson, C. D., and Kolakofsky, D. (1986). Ribosomal initiation at alternate AUGs on the Sendai virus P/C mRNA. J. Virol. 57, 684–687.

Erickson, A. H., and Blobel, G. (1979). Early events in the biosynthesis of the lysosomal enzyme cathepsin. J. Biol. Chem. 254, 11771–11774.

Evans, R. A. (1988). The steroid and thyroid hormone receptor superfamily. Science 240, 889–895.

Feagin, J. E., Abraham, J. M., and Stuart, K. (1988). Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei. Cell 53, 413–422.

Feagin, J. E., Jasmer, D. P., and Stuart, K. (1987). Developmentally regulated addition of nucleotides within the apocytochrome b transcripts in Trypanosoma brucei. Cell 49, 337–345.

Frankel, A. D., Bredt, D. S., and Pabo, C. O. (1988). Tat protein from human immunodeficiency virus forms a metal-linked dimer. Science 240, 70–73.

Galinski, M. S., Mink, M. A., Lambert, D. M., Wechsler, S. L., and Pons, M. W. (1986). Molecular cloning and sequence analysis of the human parainfluenza 3 virus mRNA encoding the P and C proteins. Virology 155, 46–60.

Giorgi, C., Blumberg, B. M., and Kolakofsky, D. (1983). Sendai virus contains overlapping genes expressed from a single mRNA. Cell 35, 829–836.

Gupta, K. C., and Kingsbury, D. W. (1985). Translational modulation in vitro of a eukaryotic viral mRNA encoding overlapping genes: ribosome scanning and potential roles of conformational changes in the P/C mRNA of Sendai virus. Biochem. Biophys. Res. Commu. 131, 91–97.

Hamaguchi, M., Yoshida, T., Nishikawa, K., Naruse, H., and Nagai, Y. (1983). Transcriptive complex of Newcastle disease virus. I. Both L and P proteins are required to constitute an active complex. Virology 128, 105–117.

Hansen, U., Tenen, D. G., Livingstone, D. M., and Sharp, P. A. (1981). T antigen repression of SV40 early transcription from two promoters. Cell 27, 603–612.

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

Hiebert, S. W., Paterson, R. G., and Lamb, R. A. (1985). Hemagglutinin-neuraminidase protein of the paramyxovirus simian virus 5; nucleotide sequence of the mRNA predicts an N-terminal membrane anchor. J. Virol. 54, 1–6.

Hull, J. D., Gilmore, R., and Lamb, R. A. (1988). Integration of a small integral membrane protein, M₂, of influenza virus into the endoplasmic reticulum: analysis of the internal signal-anchor domain of a protein with an ectoplasmic NH₂ terminus. J. Cell. Biol. 106, 1489–1498.

Jacks, T., and Varmus, H. E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science 230, 1237–1242.

Jacks, T., Townsley, K., Varmus, H. E., and Majors, J. (1987). Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumour virus gag-related polyproteins. Proc. Natl. Acad. Sci. USA 84, 4298–4302.

Johnston, M., and Dover, J. (1987). Mutations that inactivate a yeast transcriptional regulatory protein cluster in an evolutionary conserved DNA binding domain. Proc. Natl. Acad. Sci. USA 84, 2401–2405.

Koller, B., Frommn, H., Galun, E., and Edelman, M. (1987). Evidence for in vivo trans splicing of pre-mRNAs in tobacco chloroplasts. Cell 48, 111–119.

Konarska, M. M., Padgett, R. A., and Sharp, P. A. (1985). Trans splicing of mRNA precursors in vitro. Cell 42, 165–171.

Krause, M., and Hirsh, D. (1987). A trans-spliced leader sequence on actin mRNA in C. elegans. Cell 49, 753–761.

Lamb, R. A., and Lai, C.-J. (1980). Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. Cell 21, 475–485.

Lamb, R. A., and Lai, C.-J. (1982). Spliced and unspliced messenger RNAs synthesized from cloned influenza virus M DNA in an SV40 vector: expression of the influenza virus membrane protein (M1). Virology 123, 237–256.

Lamb, R. A., Etkind, P. R., and Choppin, P. W. (1978). Evidence for a ninth influenza viral polypeptide. Virology 91, 60–78.

Lamb, R. A., Lai, C.-J., and Choppin, P. W. (1981). Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. Proc. Natl. Acad. Sci. USA 78, 4170–4174.

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

Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. 65, 499–560.

McGinnes, L., McQuain, C., and Morrison, T. (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.

Moran, E., and Mathews, M. B. (1987). Multiple functional domains in the adenovirus E1A gene. Cell 48, 177–178.

Murphy, W. J., Watkins, K. P., and Agabian, N. (1986). Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. Cell 47, 517–525.

Padgett, R. A., Grabowski, P. J., Konarska, M. N., Seiler, S., and Sharp, P. A. (1986). Splicing of messenger RNA precursors. Annu. Rev. Biochem. 55, 1119–1150.

Paterson, R. G., and Lamb, R. A. (1987). Ability of the hydrophobic fusion-related external domain of a paramyxovirus F protein to act as a membrane anchor. Cell 48, 441–452.

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

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

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

Randall, R. E., Young, D. F., Goswami, K. K. A., and Russell, W. C. (1987). Isolation and characterization of monoclonal antibodies to simian virus 5 and their use in revealing antigenic differences between human, canine and simian isolates. J. Gen. Virol. 68, 2769–2780.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

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

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

Schwer, B., Visca, P., Vos, J. C., and Stunnenberg, H. G. (1987). Discontinuous transcription or RNA processing of vaccinia virus late messengers results in a 5′ poly(A) leader. Cell 50, 163–169.

Shaw, M. W., Choppin, P. W., and Lamb, R. A. (1983). A previously unrecognized influenza B virus glycoprotein from a bicistronic mRNA that also encodes the viral neuraminidase. Proc. Natl. Acad. Sci. USA 80, 4879–4883.

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

Shioda, T, Hidaka, Y., Kanda, T., Shibuta, H., Nomoto, A., and Iwasaki, K. (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. Nucl. Acids Res. 11, 7317–7330.

Solnick, D. (1985). Trans splicing of mRNA precursors. Cell 42, 157–164.

Spriggs, M. K., and Collins, P. L. (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.

Strauss, E. G., Rice, C. M., and Strauss, J. H. (1983). Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. Proc. Natl. Acad. Sci. USA 80, 5271–5275.

Sutton, R. E., and Bothroyd, J. C. (1986). Evidence for trans splicing in trypanosomes. Cell 47, 527–535.

Tabor, S. (1987). Sequenase™: step-by-step protocols for DNA sequencing with Sequenase™. United States Biochemical Corporation. Cleveland, OH.

Takeuchi. K., Hishiyama, M., Yamada, A., and Sugiwa, A. (1988). Molecular cloning and sequence analysis of mumps virus gene encoding the P protein: mumps virus P gene is monocistronic. J. Gen. Virol., in press.

Varmus, H., (1988). Retroviruses. Science 240, 1427–1435.

Yoshinaka, Y., Katoh, I., Copeland, T. D., and Oroszlan, S. (1985). Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA 82, 1618–1622.

Zoller, M. J., and Smith, M. (1982). Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment. Nucl. Acids Res. 10, 6487–6500.