The P Gene of Human Parainfluenza Virus Type 1 Encodes P and C Proteins but Not a Cysteine-Rich V Protein

YUMIKO MATSUOKA,¹ JOSEPH CURRAN,² THIERRY PELET,² DANIEL KOLAKOFSKY,²
RANJIT RAY,¹ AND RICHARD W. COMPANS^{3*}

Secretech Inc., Birmingham, Alabama 35205¹; Department of Microbiology, University of Geneva Medical School, CH-1211 Geneva 4, Switzerland²; and Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294³

Received 15 January 1991/Accepted 16 March 1991

The nucleotide sequence of the P gene of human parainfluenza virus type 1 (PIV1) was determined from cloned cDNA copies of the mRNA. By analogy with the gene organization of Sendai virus, two open reading frames in the mRNA sense of the gene were identified as coding sequences for the P protein (568 amino acids with an estimated molecular weight of 64,655) and the C protein (204 amino acids with an estimated molecular weight of 24,108). Comparison of the deduced amino acid sequences of the P and C proteins of PIV1 with those of Sendai virus showed a high degree of homology. However, a sequence for the cysteine-rich V protein, which was considered a common feature of other paramyxoviruses, was interrupted by the presence of multiple stop codons. The sequence analysis of three P-gene-specific cDNA clones generated from genomic RNA by polymerase chain reaction and one additional clone generated from mRNA confirmed that the coding sequence for the cysteine-rich region is silent in the PIV1 gene and thus is not translated into protein. Two potential editing sites with the consensus sequence 3'UUYUCCC were found in the PIV1 P gene at positions 564 to 570 and 1430 to 1436. However, examination of the PIV1 mRNA population by a primer extension method indicated that neither of these sites is utilized. These results indicate that the PIV1 P gene has a coding strategy different from those of other paramyxovirus P genes.

The P genes of paramyxoviruses produce multiple proteins from alternate reading frames, and various mechanisms are involved in the synthesis of these proteins. The P genes of simian virus 5 (27), measles virus (2), Sendai virus (SEN) (29), human parainfluenza virus type 2 (PIV2) (21, 26), and mumps virus (8) were found to produce more than one mRNA species with different sizes by inserting different numbers of G residues at a precise site in the gene. mRNA species which are exact transcripts of the P gene of SEN and measles virus produce P proteins, whereas edited mRNA species with one additional G residue are translated into V proteins (2, 29). In contrast, P proteins of simian virus 5, PIV2, and mumps virus are synthesized from mRNA species in which two G residues are inserted, and faithfully transcribed mRNAs encode V proteins. The P protein, a highly phosphorylated protein, is a component of viral nucleocapsids, and both P and L proteins are required for RNA transcriptase activity (14). P and V proteins initiate from the same AUG start codon and share the N-terminal amino acid sequence upstream of the G residue insertion site. The V protein contains a cluster of cysteine residues in the C-terminal portion that is not shared with P and may possess metal-binding activity (27). In addition to the mRNA editing mechanism, several nonstructural proteins, referred to as the C proteins, are also synthesized from the third reading frame of the P gene from different initiation sites, which include an ACG start codon (6, 7, 13). A small protein, designated X, is also synthesized from the same open

reading frame (ORF) as the P protein of SEN, initiating at an internal methionine codon which is more than 1,500 nucleotides (nt) distant from the 5' end of mRNA (5).

PIV1 is closely related to SEN (4, 12, 15, 31). The nucleotide sequences of the fusion (F) (20), hemagglutinin-neuraminidase (HN) (11, 18), and nucleoprotein (NP) genes (17) of PIV1 have been determined and have shown a striking homology with the corresponding genes of SEN. In this study we investigated whether the P gene of PIV1 possesses an expression strategy similar to that found in the SEN P gene.

Sequence analysis and organization of the P gene. PIV1 (strain C39) was obtained from J. Kasel, Baylor College of Medicine, Houston, Tex. The virus was propagated in Vero cells and purified as described previously (18). The construction and identification of PIV1-specific cDNA clones have been described previously (18). PIV1-specific clones were characterized by hybridization, and those analyzed in this study did not hybridize with riboprobes made from SEN HN and F clones. Clones containing PIV1-specific genes derived from mRNA were obtained, and their nucleotide sequences were partially analyzed. Comparison of these partial sequences with the SEN P gene indicated that two clones, 11-26 and 10-56, represented the PIV1 P gene. The nucleotide sequences of both strands of the entire inserts of the two clones were then determined. The 11-26 clone was found to contain 1,873 bp, including a 10-nt stretch of poly(A) at the putative 3' end. The sequence UAAG preceding the poly(A) tail was identical to the consensus termination sequence found in the SEN genes (10). The 10-56 clone was 82 bp shorter than 11-26 at the 5' end but otherwise was identical.

* Corresponding author.


```

PIV3  GHRREHSIYREGDYIITESWCNPICSKIRPVPRQESCVCGECPKQCGYCIEDRLSSRSNDWSINGS  412
      ::::: ::  :: ::::: :: :: :: :: :: :: :: :: :: :: :: ::
SEN   GHRREHIYERDGYIVDESWCNPFVCSRIRVPRRELVCVKCTCPKVKLCRDDIQCMRPDPFCREIFRS  384
      :  ::  :: :::::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
PIV1  *TGNQQHISR*DGQIVSKSWCNEITT*I*II*KCKSCIC*AHVKICKLQRNDI*SLWYAYIS*KIT*E

```

FIG. 2. Alignment of the V protein regions of SEN, PIV1, and PIV3. The cysteine-rich region of SEN and PIV3 (bovine) V proteins and the corresponding sequence of PIV1 are compared. Identical amino acids are indicated by double dots, cysteine residues conserved among the three viruses are marked by closed circles, and those conserved between SEN and PIV3 are marked by open circles above the sequences. Stop codons in the PIV1 sequence are indicated by asterisks.

have been detectable. When the same method is applied to SEN, measles virus, and bPIV3 P gene mRNAs, 25 to 65% of the mRNAs are found to have one or more insertions (23). It would therefore appear that the PIV1 P gene expresses only the P and C proteins and that the presence of the sequence 3'UUYUCCC is insufficient for polymerase stuttering, presumably because the PIV1 polymerase does not pause at these sites.

Implications of the PIV1 P gene sequence. When the predicted sequence of the PIV1 P protein was aligned with that of other paramyxoviruses, it was found to be most closely related to SEN and then to bPIV3. This was expected, since the same relationship had previously been found in comparisons of the F, HN, and NP proteins as well as in serological studies (4, 12, 15, 17, 18, 20, 31). The PIV1 and SEN P proteins are both 568 residues in length and can be aligned without gaps (Fig. 4), and 53% of the amino acids have been strictly conserved between them. The homology is not uniformly distributed, however, but is strongest in two large blocks. Residues 1 to 232 and 324 to 568 (regions 1 and 3) contain 52 and 69% identical residues, respectively,

whereas residues 233 to 323 (region 2) contain only 11% homology. The same pattern is found in comparing PIV1 and bPIV3, except that the homology is much weaker (regions 1 to 3 have 23, 8, and 33% identical residues, respectively). Until the PIV1 P sequence was deduced, the SEN-bPIV3 alignment was too ambiguous to draw structural conclusions. The three-way alignment in Fig. 4, however, is quite clear. On the basis of the pattern of conservation, these P proteins are composed of two separate domains. The more highly conserved C-terminal domain, or region 3, has been shown to contain the sites with which P binds to the NP protein of the viral nucleocapsids (16, 24). Region 1, on the other hand, contains the sites where the protein is modified by phosphorylation (30). Similar N- and C-terminal domains are also found in the rhabdovirus vesicular stomatitis virus P protein (called NS) (3, 9).

The SEN and bPIV3 P genes also contain two additional ORFs, designated C and V, which overlap that of P in the +1 and -1 frames, respectively. The C ORF is found at the N terminus of P and is expressed by independent ribosomal initiation (6). The V ORF is found near the middle of the P

FIG. 3. Examination of mRNA insertions at two possible sites. The method is schematized at the top for the predicted site at position 564 to 570. For the site at position 1430 to 1436, the primer 3'GTIATTCCAGTTTTC (complementary to position 1442 to 1457) was used, and extension was terminated with ddATP opposite the U in the mRNA at position 1426. The reactions were carried out as described previously (23), except that 20 μ g of CsCl pellet RNA was used and the deoxynucleoside triphosphate and ddATP concentrations were 200 μ M. The top line shows the relevant region of the mRNA with the GGG (position 568 to 570) within which insertions would occur printed in bold and underlined. The bottom line shows the extended primer terminated with dideoxyadenosine, with the primer underlined and printed in bold. The bottom panel shows a 12% sequencing gel of the extended primers. T7, mRNAs made in vitro with T7 polymerase from plasmid DNA; + and -, CsCl pellet RNA from infected and uninfected BHK or CV1 cells, respectively. The numbers on the side refer to the length of the extended primers (in nucleotides).

SEN	MDQDAFILKEDSEVEREAPGGRESLSDVIGFLDAVLSEPTDIGGDRSWLHNTINTPQGPSAHRKSEGEVSTPS	-78	
PIV1	MDQDAFFERPEAGEAPRKQESLGVIGLLDVLSYKPTTEIGEDRSWLHGIIDNPEENKPSCKADDNNKDRAISTP	-78	52%
PIV3	MENNAKDNQIMDSWEEGSGDKSSDISSALDIEFILSTDSQENTADSNEVN	-72	23%
SEN	TQDNRSGEESRVSGRTSKPEAAHAGNLDKQNIHRAFGGRTGTNSVSDLDGDDSGILENPPNERGVPRSGIEDEN	-155	
PIV1	TQDHRSGEESGISRRTSESKTETHARLLDQOSIHRASRGTSFNPPLPENMGNERNTRIDEDSPNERRHQRSVLTDED	-155	52%
PIV3	KENSGSVNENRQLGASHERATETKNRNVNQETIQGGNRGRSSSDSRAEIMVTRGISRSPDPNNGTQIQESIDYNEV	-149	23%
SEN	REMAAHPDKRGEDQAEGLPEEVGGTSLPDEGEAGASNNGRSMEPGSSHSARVTGVLPISPELEEAVALRRNKRRT	-232	
PIV1	RKMAEDSNKREEDQVEGFPEEIRRTPLSDGESRTNNGRSMETSSSTHSTRITDVIINPSPELEDAVLQRNKRRT	-232	52%
PIV3	GEMDKDSAKRENMRQSKDVPVKVSRSDAIPPTKQDGNDDGRSMESISTFDSGYTSIVTAATLDDEEELMKNTRPKR	-226	23%
SEN	NSGSKPLTPATVPGTRSPPLNRYNSTGSPGKPPSTQDEHINSQDTPAVRVKDRKPPIGTRSVSDCPANGRPIHPGLESSTKKGIGENTS	-323	
PIV1	IIRNQTRSERQNSLHKSTSENSNLEDHNTKTSKGLPPKNEESAATPKNNHNRKTKYTMNNANNTKSPPTPEHDTTANEETSNT	-323	11%
PIV3	GSKRNQTLKISTTTGSTRPQSGSQGKRITSWNILNSESGRTESTSQNSQIPTSGKNSNTVGPGRITLESRIKTQKTGKEREDTEESTR	-353	8%
	YQSTPQEDDKGIIKKVGKPEDTNKQSPILDYELNSK		
SEN	SMKEMATLLTSLGVIQSAQEFSSRDASYVFAARRAKSANYAEMTFNVCGLIISAEEKSSARKVDENKQLLKQIQESV	-400	
PIV1	SVDEMAKLLVSLGVMSKQHEFELSRASHVFAKRLKSANYKEMTFNLGMLISVEKSLNKKVEENRTLLKQIQEEI	-400	69%
PIV3	FTERAITLLQNLGVIQSAAKLDLYQDKRVVCVANVLNADTASKIDFLAGLMIGVSMHDHTKLNQIQNEILSKTDL	-430	33%
SEN	ESFRDIYKRFSEYQKEQNSLLMSNLSTLHIITDRGGKTDNTDSLTRSPSVFAKSKENKTKATRFDPSPMETLEDKMKYPDL	-480	
PIV1	NSSRDHLKRFSEYQKEQNSLMMANLSTLHIITDRGGKTDGPDSTTRSPSVFTKGDKNVKKTRFDPSMEALGGQEFKPD	-480	69%
PIV3	KKMDESHRLIENQEQLSLITSLISNLKIMTERGGKKDQENSGRTPMIKTKAKEEIKKVRFDPLMETQGIENIPDL	-510	33%
SEN	IREDEFDEIRNPVYQERDTEPRASNARLLPSKEKPTMHSRLRLVIESSPLSRAEKAAYVKSLSKCKTDQEVKAMVLEEDIESLTN	-568	
PIV1	IREDELDDIKNPLENNNDLQASNARLNPSTEKHTLSLKLVIENSPLSRVEKKAYIKSLYKCRTNQEVKNMVEEDIDSLTN	-568	69%
PIV3	YRSIEKTPENDIQIKSDINRSNDESNAITRVLPKRTSNTMRLIIINNSNLSSRAKQSYINELKCKSDEEVSELMDFNEDVSS--Q	-596	33%

FIG. 4. Amino acid sequence alignment of P proteins of SEN, PIV1, and PIV3 (bovine). Proteins were aligned in pairs by an alignment program. Gaps were inserted to maximize the number of matches between PIV1 and PIV3. Identical amino acids are indicated by double dots and the percentages of identity for each region (region 1, 1 to 232; region 2, 234 to 323; and region 3, 324 to 568 for SEN and PIV1 and region 1, 1 to 226; region 2, 227 to 353; and region 3, 354 to 596 for PIV3) are shown on the right margin. Overlapping coding regions for the C and V proteins of SEN are indicated by the letters c and v, respectively, above the SEN sequence. The insertion sites for extra G residues in SEN and PIV3 are underlined and also marked by asterisks. The sequence KKG representing the nucleotide sequence AAA AAA GGG is maintained in both SEN and PIV3.

gene and is expressed as a fusion protein together with the N-terminal half of P, from a separate mRNA, referred to as V/C. The V/C mRNA differs from the P/C mRNA by the insertion of a single G within a short run of G's because of polymerase stuttering at this site during mRNA synthesis (28, 29). Both the SEN and bPIV3 editing sites are found within the least conserved central region 2, but at opposite ends. The SEN site is at the C terminus of region 2, whereas that of bPIV3 is at the N terminus, within an additional 36 residues not found in SEN or PIV1. The other proposed editing site of bPIV3, near the SEN site and 100 residues downstream, does not function in bPIV3 (23).

Remarkably, inspection of the PIV1 P gene shows that it appears not to contain a functional V ORF. The remnants of the V ORF are easily detected, but this frame has been closed by no less than nine stop codons within these 68 residues. The minimum editing sequence of other paramyxovirus P genes (3'UUYUCCC) is also absent from the region upstream of the remnants of the V ORF. This sequence is found at two other locations in the PIV1 P gene, but we have been unable to find evidence of insertions in the mRNA at

these sites. It therefore appears that the PIV1 P gene neither expresses a V protein nor edits its mRNA for other purposes. In contrast to the PIV1 V region, the PIV1 C ORF has been well conserved with that of SEN. The nucleotide sequence containing the P and C protein AUGs has been maintained almost exactly between PIV1 and SEN, within the common sequence 5'CCGNAUGGAUCARGAUGCCUUC, and aa 77 to 204 are 88% identical. However, the upstream ACG which starts the SEN C' protein is not present in PIV1 (although it is present in bPIV3) and the downstream AUG's which start the SEN Y1 and Y2 proteins are absent as well, as in bPIV3. Since the V ORF is significantly more conserved than either the P or C ORFs when all members of the paramyxoviruses and morbilliviruses are considered (27), it remains possible that PIV1 expresses a V protein from another region of its genome which has not yet been sequenced. It is thus presently unclear whether PIV1 is the only member of these viruses (except for respiratory syncytial virus) which does not express a V protein. By comparison with most other paramyxoviruses which possess a V protein, PIV1 replicates

to lower levels and exhibits little or no detectable cytopathic effect on host cells. PIV1 replication is therefore usually detected by hemadsorption in cell cultures exhibiting no cytopathic effects. If PIV1 does not express a V protein, it will be of interest to determine whether the presence of a functional V protein may play a role in facilitating higher levels of virus replication. By using approaches developed with measles virus, it may now be possible to determine whether the V and C function are essential for viral replication (1).

Nucleotide sequence accession number. The sequence data in this paper have been deposited with the EMBL and GenBank libraries under accession no. M37792.

This study was supported by contract N43-AI-95019 and grant AI12680 from the National Institute of Allergy and Infectious Diseases and by the Swiss National Science Fund.

REFERENCES

- Ballart, I., D. Eschle, R. Cattaneo, A. Schmid, M. Metzler, J. Chan, S. Pifko-Hirst, S. A. Udem, and M. A. Billeter. 1990. Infectious measles virus from cloned cDNA. *EMBO J.* **9**:379-384.
- Cattaneo, R., K. Kaelin, K. Bacsko, and M. A. Billeter. 1989. Measles virus editing provides an additional cysteine-rich protein. *Cell* **56**:759-764.
- Chattopadhyay, D., and A. K. Banerjee. 1987. Two separate domains within vesicular stomatitis virus phosphoprotein support transcription when added in trans. *Proc. Natl. Acad. Sci. USA* **84**:8932-8936.
- Cook, M. K., B. E. Andrew, H. H. Fox, H. C. Turner, W. D. James, and R. M. Chanock. 1959. Antigenic relationships among the 'newer' myxoviruses (parainfluenza). *Am. J. Hyg.* **69**:250-264.
- Curran, J., and D. Kolakofsky. 1988. Scanning independent ribosomal initiation of the Sendai virus X protein. *EMBO J.* **7**:2869-2874.
- Curran, J., and D. Kolakofsky. 1989. Scanning independent ribosomal initiation of the Sendai virus Y proteins *in vitro* and *in vivo*. *EMBO J.* **8**:521-526.
- 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.
- Elliott, R. P., M. A. Yeo, E. J. B. Afzel, J. A. Curran, and B. K. Rima. 1990. Strain-variable editing during transcription of the P gene of mumps virus may lead to the generation of non-structural proteins NS1 (V) and NS2. *J. Gen. Virol.* **71**:1555-1560.
- Emerson, S. U., and M. Schubert. 1987. Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA* **84**:5655-5659.
- Giorgi, C., B. M. Blumberg, and D. Kolakofsky. 1983. Sendai virus contains overlapping genes expressed from a single mRNA. *Cell* **35**:829-836.
- Gorman, W. L., D. S. Gill, R. A. Scroggs, and A. Portner. 1990. The hemagglutinin-neuraminidase glycoproteins of human parainfluenza virus type 1 and Sendai virus have high structure-function similarity with limited antigenic cross-reactivity. *Virology* **175**:211-221.
- Goswami, K. K. A., and W. C. Russell. 1983. Monoclonal antibodies against human parainfluenza type 3 and against SV5 virus: preparation and preliminary characterization. *J. Gen. Virol.* **64**:1663-1672.
- Gupta, K. C., and S. Patwardhan. 1988. ACG, the initiator codon for a Sendai virus protein. *J. Biol. Chem.* **263**:8553-8556.
- Hamaguchi, M., T. Yoshida, K. Nishikawa, H. Naruse, and Y. Nagai. 1983. Transcriptional complex of Newcastle disease virus. I. Both L and P proteins are required to constitute an active complex. *Virology* **128**:105-117.
- Ito, Y., M. Tsurudome, M. Hishiyama, and A. Yamada. 1987. Immunological interrelationships among human and non-human paramyxoviruses revealed by immunoprecipitation. *J. Gen. Virol.* **68**:1289-1297.
- Kolakofsky, D., S. Vidal, and J. Curran. Paramyxovirus RNA synthesis and P gene expression. In D. W. Kingsbury (ed.), *The paramyxoviruses*, in press. Plenum Press, New York.
- Matsuoka, Y., and R. Ray. 1991. Sequence analysis and expression of the human parainfluenza type 1 virus nucleoprotein gene. *Virology* **181**:403-407.
- Matsuoka, Y., R. Ray, and R. W. Compans. 1990. Sequence of the hemagglutinin-neuraminidase gene of human parainfluenza virus type 1. *Virus Res.* **16**:107-114.
- Maxam, A., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
- Merson, J. R., R. A. Hull, M. K. Estes, and J. A. Kasel. 1988. Molecular cloning and sequence determination of the fusion protein gene of human parainfluenza virus type 1. *Virology* **167**:97-105.
- Ohgimoto, S., H. Bando, M. Kawano, K. Okamoto, K. Kondo, M. Tsurudome, M. Nishino, and Y. Ito. 1990. Sequence analysis of P gene human parainfluenza type 2 virus: P and cysteine-rich proteins are translated by two mRNAs that differ by two nontemplated G residues. *Virology* **177**:116-123.
- Paterson, R. G., and R. A. Lamb. 1990. RNA editing by G-nucleotide insertion in mumps virus P-gene mRNA transcripts. *J. Virol.* **64**:4137-4145.
- Pelet, T., J. Curran, and D. Kolakofsky. 1991. The P gene of bovine parainfluenza virus 3 expresses all three reading frames from a single mRNA editing site. *EMBO J.* **10**:443-448.
- Sakai, Y., S. Suzu, T. Shioda, and H. Shibuta. 1987. Nucleotide sequence of the bovine parainfluenza 3 virus genome: its 3' end and the genes of NP, P, C and M proteins. *Nucleic Acids Res.* **15**:2928-2944.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
- Southern, J. A., B. Precious, and R. E. Randall. 1990. Two nontemplated nucleotide additions are required to generate the P mRNA of parainfluenza virus type 2 since the RNA genome encodes protein V. *Virology* **177**:388-390.
- 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.
- Vidal, S., J. Curran, and D. Kolakofsky. 1990. A stuttering model for paramyxovirus P mRNA editing. *EMBO J.* **9**:2017-2022.
- Vidal, S., J. Curran, and D. Kolakofsky. 1990. Editing of the Sendai virus P/C mRNA by G insertion occurs during mRNA synthesis via a virus-encoded activity. *J. Virol.* **64**:239-246.
- Vidal, S., J. Curran, C. Orvell, and D. Kolakofsky. 1988. Mapping of monoclonal antibodies to the Sendai virus P protein and the location of its phosphates. *J. Virol.* **62**:2200-2203.
- Yoshida, T., Y. Nakamura, T. Toyoda, T. Nishikawa, M. Hamaguchi, and Y. Nagai. 1988. The polypeptides of human parainfluenza 1 (HA2) virus and their antigenic relationships to those of Sendai virus (HVJ). *Microbiol. Immunol.* **29**:577-582.