JOURNAL OF VIROLOGY, Mar. 1992, p. 1765–1768
0022-538X/92/031765-04$02.00/0

Vol. 66, No. 3

The Parainfluenza Virus Type 1 P/C Gene Uses a Very Efficient GUG Codon To Start Its C′ Protein†

RONALD BOECK,[1] JOSEPH CURRAN,[1] YUMI MATSUOKA,[2] RICHARD COMPANS,[3] AND DANIEL KOLAKOFSKY[1]*

Department of Genetics and Microbiology, University of Geneva School of Medicine, 9, Avenue de Champel, CH-1211 Geneva, Switzerland[1]; Secretech, Inc., Birmingham, Alabama 35205[2]; and Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294[3]

Received 23 July 1991/Accepted 19 November 1991

Parainfluenza virus type 1 (PIV1) and Sendai virus (SEN) are very closely related, but the PIV1 P/C gene does not contain the ACG codon which initiates the SEN C′ protein. Nevertheless, a protein corresponding to the PIV1 C′ protein was observed both in vivo and in vitro. The initiation site of this protein maps upstream of the PIV1 C protein AUG in a region that does not contain an AUG codon. We have used site-directed mutagenesis to demonstrate that the PIV1 C′ protein initiates from a GUG codon, four codons upstream of where the ACG is found in SEN. Remarkably, this GUG appears to initiate in vivo almost as frequently as AUG in the same context. However, whereas GUG permits downstream expression of the P and C proteins, AUG in this context does not. The conservation of an upstream non-AUG initiation codon for C′ among PIV1 and SEN suggests that it is important for virus replication, even though some paramyxoviruses express only the C protein and others have no C open reading frame at all.

Human parainfluenza virus type 1 (PIV1) and Sendai virus (SEN) are very closely related paramyxoviruses, only slightly less closely related than human and bovine PIV3. Although originally isolated from young infants (16), SEN is also endemic in mice and is sometimes referred to as murine PIV1. Together, these four viruses form one related group within the paramyxovirus genus, whereas simian virus 5, PIV2, PIV4, and mumps virus form another group (14).

The sequences of the PIV1 NP (17, 19), P/C (18), F (21), and HN genes (10, 20) have recently been determined. The P/C gene was of particular interest, because this gene of SEN expresses multiple proteins by using two overlapping open reading frames (ORFs), alternate initiation codons, and edited mRNAs (5, 6, 27) (Fig. 1). The PIV1 and SEN P/C genes are very similar. In both cases, the first AUG starts the P protein of 568 residues and the second AUG starts the C protein of 204 residues in the +1 frame relative to P. However, the PIV1 P/C gene has several distinctive features. For example, an ORF which codes for a protein designated V, which was thought to be a common feature of paramyxoviruses and morbilliviruses, cannot be expressed in PIV1 because it is interrupted by no fewer than nine stop codons, and a site for G insertion or editing, which is necessary for production of the V mRNA, is absent as well (18). In addition, whereas the SEN P/C gene uses first an ACG codon and then three successive AUG codons to start a nested set of four related C proteins, designated C′, C, Y1, and Y2 (5, 6, 11), the PIV1 P/C gene has maintained only the AUG for the C protein. Nevertheless, when a plasmid-derived PIV1 P/C gene mRNA was translated in vitro, two C proteins were clearly identified. In this study, we have used site-directed mutagenesis to investigate the mechanism of initiation of the second C protein, designated C′, of PIV1.

PIV1 C′ protein initiates from a GUG codon. Two PIV1 P/C clones, 11-26 and 10-56, were originally sequenced (18). They were found to be identical, except that 11-26 was 82 nucleotides (nt) longer at its 5′ end and contained 74 nt before the P protein AUG start codon. The 5′ end of clone 10-56 began between the P and C protein AUGs, which are very well conserved between PIV1 and SEN, within the common sequence 5′CCGNAUGGAUCARGAUGCC. When mRNA derived from clone 11-26 (designated wt in Fig. 2A) was translated in a wheat germ extract, three major products were seen (Fig. 2): the putative P protein with an electrophoretic mobility slightly less than that of its SEN counterpart, and two bands which migrated in the range of the SEN C′ and C proteins (comparisons not shown). Immunoprecipitation with specific antibodies confirmed the identities of the in vitro-made proteins (Fig. 2B). The same two bands are found on Western blots (immunoblots) of PIV1-infected cell extracts developed with polyclonal antiserum to the SEN C proteins (not shown). We expected this antiserum to cross-react with the PIV1 C proteins, as residues 77 to 204 of these sequences are 88% identical. Both in vivo during natural infection and in vitro, C′ is the predominant C protein in PIV1, whereas the opposite is the case for SEN (5, 6).

When clone 10-56 mRNA was translated in vitro, C protein (but not P protein) was made as expected but there was no evidence that C′ was made (not shown). These results indicated that the two PIV1 C proteins were not due to partial posttranslational modification of a single protein. Rather, the PIV1 gene also appeared to code for C′, initiated upstream within the 82 nt missing from the shorter mRNA. Inspection of this sequence showed no suitable ACG codon as found in the SEN sequence (Fig. 1). However, there was a GUG codon in an excellent context for ribosomal initiation (ACAGUGG), four codons upstream from where the ACG is found in SEN. To examine whether this GUG was responsible for the C′ protein initiation, we used site-directed mutagenesis to change it to GCG (C′ GCG, which has never acted as an initiation codon in the SEN P/C gene [6]) as well

* Corresponding author.

† Hiroshi Shibuta of the Institute of Medical Sciences, University of Tokyo, who contributed greatly to our understanding of paramyxoviruses, died on 20 November 1991. This article is dedicated to his memory.

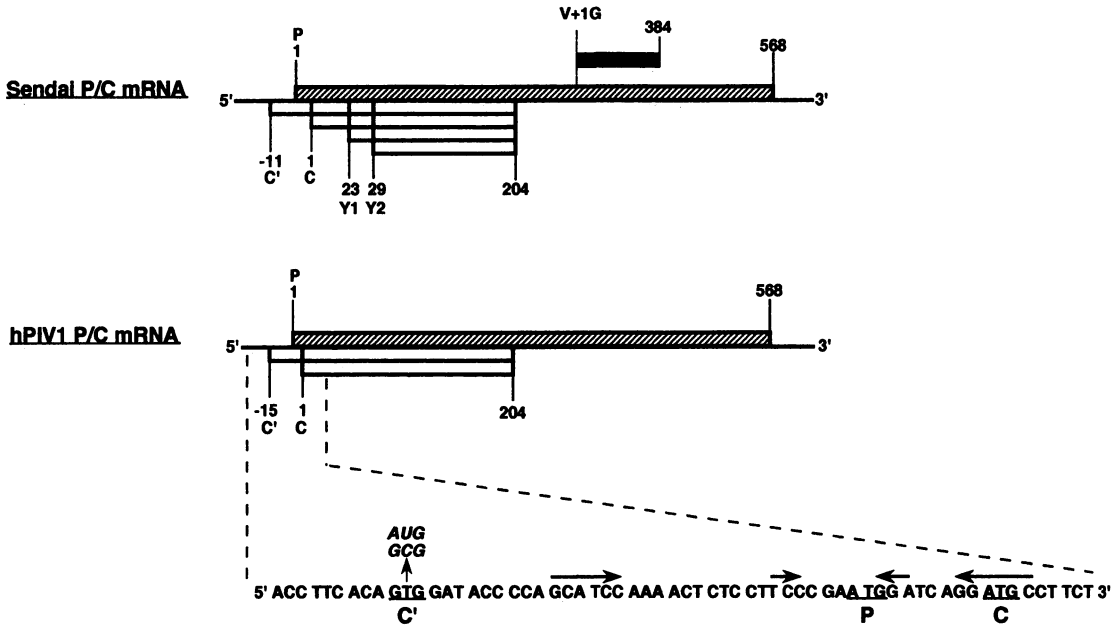

FIG. 1. Schematic representation of the P/C gene mRNAs of SEN and PIV1. The mRNAs are represented as horizontal lines whose 5′ and 3′ ends are indicated. Hatched boxes (above the line) refer to the P ORFs; open boxes (below the line) refer to the various C ORFs. Each C ORF attached directly to the mRNA line represents an independently initiated protein. The V ORF of SEN (black box) is shown not attached to the mRNA line, as it is translated only in edited mRNAs which contain the +1G insertion. The initiation sites are indicated by letters, and the numbers refer to amino acid positions; for the C proteins, the AUG starting the C protein is set as number 1. The sequence surrounding the initiation sites of the PIV1 C′, P, and C proteins is shown below its mRNA. The initiation codons are underlined, and the two point mutations changing the GUG start codon are indicated. The horizontal arrows above the sequence indicate a possible stem-loop structure downstream of the C′ GUG, which is predicted to have a free energy of −14 kcal/mol (ca. −59 kJ/mol).

as to AUG (C′ AUG). When the plasmid-derived mRNAs were translated and their products were compared with those of the wild-type construct (Fig. 2), the results were entirely consistent with this hypothesis. C′ AUG mRNA made essentially only C′. The bands just above C′ and C (unmarked in Fig. 2) are likely to be posttranslationally modified forms of these proteins, as a similar situation occurs for SEN (5, 6). Initiation from the downstream AUGs for P and C is extremely inefficient in C′ AUG, presumably because the new upstream AUG was nonleaky for scanning ribosomes. The C′ GCG mRNA, on the other hand, made no C′ protein but did make enhanced amounts of P and C, presumably because the GUG-to-GCG change allowed more ribosomes to scan downstream. Immunoprecipitation with specific antibodies again confirmed that these translation products contained the appropriate sequences (Fig. 2B).

To examine the start codon usage in vivo, the wild-type and mutant constructs were transferred to the cytomegalovirus-promoted vector pSC (12a), expressed in COS cells, and protein levels were estimated by immunoprecipitation (Fig. 3 and Table 1). The results were very similar to those found in vitro. C′ AUG made slightly more C′ than wild-type mRNA but virtually eliminated the downstream P and C initiation, whereas C′ GCG eliminated C′ synthesis and increased P and C synthesis. Clearly, the PIV1 C′ protein initiates from this GUG codon and contains 15 residues in addition to the 204-amino-acid C protein.

Use of non-AUG start codons. Non-AUG ribosomal initiation codons are still relatively rare, but their list is steadily increasing. For eucaryotes, ACG is used in both adeno-associated virus (2) and SEN (6, 11) as well as in a form of phosphoribosylpyrophosphate synthetase (26); CUG is used in murine retroviruses (murine leukemia virus) (22) and the cellular proto-oncogenes *myc* (12), *int-2* (1), and *pim-1* (24) as well as in basic fibroblast growth factor (8, 23) and the ltk receptor tyrosine kinase (3); and GUG has recently been reported to initiate drosophila choline acetyltransferase (25) and is now demonstrated to be utilized in PIV1.

Kozak (15) has examined six potential non-AUG initiation codons in vivo by using them to replace the AUG of model mRNAs such as chloramphenicol acetyltransferase or preproinsulin. Of these, only GUG worked to any extent, but only at 3 to 5% of the level of AUG. She has argued that initiation at non-AUG codons in vivo is necessarily inefficient, as is the case with AUG start codons in poor contexts, and that results in vitro showing relatively efficient use of these codons are probably artifactual. Since then, however, Saris et al. (24) have reported that the upstream CUG in the *pim-1* mRNA is used at about 50% of the efficiency of AUG in the same context in COS cells, and the results in Fig. 3 and Table 1 show even higher efficiencies for the PIV1 GUG (also in COS cells). To determine whether this remarkably high efficiency was peculiar to the expression system used, the wild-type and C′ AUG pGEM constructs of Fig. 2 were also expressed by transfection in BHK cells, and T7 polymerase was provided with a recombinant vaccinia virus (9). When the protein gel of the immunoprecipitated products was quantitated (Table 1, experiment 2), we found essentially the same results. Replacing this GUG with AUG led to only a marginal increase (<10%) in C′ expression but nevertheless again effectively eliminated downstream expression of P and C. The high efficiency of this non-AUG start site is thus the same, even when estimated in very different expression systems.

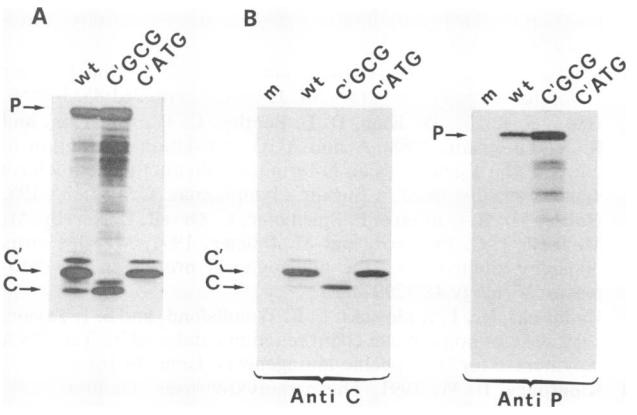

FIG. 2. Translation of wild-type and mutant mRNAs in vitro. Polymerase chain reaction was used to introduce base changes into the pGEM4 PIV1 P/C wild-type clone (13). Briefly, the upstream 25-mer message-sense primer was the mutagenic oligonucleotide and corresponded to nt 36 to 52 of the sequence (18), except for the single base change plus a 5′ dangling *Sac*I restriction site. The downstream primer was a 14-mer corresponding to nt 123 to 136 (antisense strand) 10 nt downstream of a naturally occurring *Bam*HI site. The fragments amplified by polymerase chain reaction were digested with *Sac*I and *Bam*HI and used to replace the same fragment of pGEM4 PIV1 P/C to create pGEM C′ AUG and GCG. (A) pGEM PIV1 P/C wild type (wt), C′ AUG, and C′ GCG were transcribed in vitro, and an equal amount of mRNA (50 ng) was translated in a wheat germ extract in the presence of [^{35}S]methionine. One-fifth of each reaction was analyzed directly on a 12.5% polyacrylamide gel. (B) Immunoprecipitation of the proteins expressed by the wild-type (wt), C′ GCG, and C′ ATG mRNAs. Lane m, control reaction in which no mRNA was added to the wheat germ extract. One-fifth of each reaction was incubated with either a SEN C and C′ antiserum (Anti C) or a PIV1 antiserum to whole virus (Anti P). Antigen-antibody complexes were recovered and resolved on a 15% (Anti C) or 12.5% (Anti P) polyacrylamide gel.

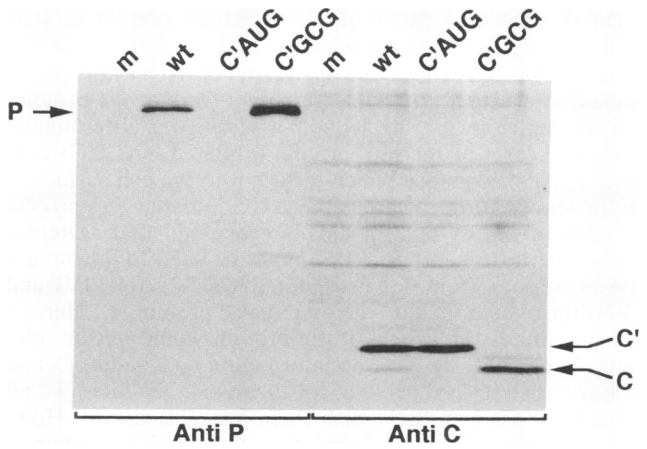

FIG. 3. Expression of wild-type and mutant P/C genes in vivo. The P/C genes of the wild-type (wt), C′ AUG, and C′ GCG pGEM constructs were subcloned into pSC (12a), and 5-μg samples of these plasmid DNAs were transfected into parallel 5-cm dishes of COS cells (4). As a control, one culture was mock transfected (lanes m). All the cultures were labelled with 150 μCi of [^{35}S]methionine for 1 h at 60 h posttransfection. Cells were disrupted by sonication in radioimmunoprecipitation assay buffer containing 2% sodium dodecyl sulfate and 1% 2-mercaptoethanol, and one-half of each sample was immunoprecipitated with either polyclonal antiserum to the SEN C proteins or a mixture of two monoclonal antibodies to the PIV1 P protein.

In contrast, when the C′ ACG start codon in SEN was changed to AUG, C′ expression increased sevenfold (5), and replacing the ACG of SEN with GUG also led to a modest (twofold) improvement in the efficiency of C′ expression (20a). GUG then appears more efficient than ACG for ribosomal initiation, a finding consistent with previous studies (15). mRNAs designed to use non-AUG start codons at high efficiency, moreover, may contain additional elements which are important. Kozak (15a) has shown that the introduction of a modest secondary structure (a stem-loop with a free energy of −19 kcal/mol [ca. −79 kJ/mol]), when judiciously positioned downstream, can significantly increase expression from non-AUG codons. There is, in fact, a modest secondary structure (−14 kcal/mol [ca. −59 kJ/mol], indicated in Fig. 1) which begins 10 nt downstream of the PIV1 GUG, and this may also contribute to the efficiency of utilization of this GUG. However, we note that while these putative elements may allow this GUG to initiate almost as frequently as an AUG in the same context, the PIV1 GUG is much less efficient at preventing ribosomes from scanning downstream. Similar results were also noted by Mehdi et al. (20a) when they replaced the ACG in SEN with CUG and GUG. We have no plausible explanation of this apparent uncoupling of initiation efficiency and leaky scanning within the present view of ribosomal initiation. Moreover, it is unclear whether this apparent uncoupling is peculiar to non-AUG start sites.

Paramyxovirus C protein(s). Little is presently known about the paramyxovirus C proteins, except that they are uniformly basic and very underrepresented in virions (7). Moreover, in some viruses only a single protein is apparently generated from this ORF, whereas other viruses have no C ORF at all (14). The C proteins therefore cannot provide an essential function that is common to all members of this family. Nevertheless, the fact that many of these viruses have conserved this ORF suggests that it plays an important role when it is present. Similarly, the finding that the PIV1 P/C gene has not conserved the ACG which initiates C′ in SEN but uses a GUG codon four codons upstream instead suggests that the presence of a non-AUG start codon in this gene is important in some respect for viral replication.

TABLE 1. Relative synthesis from mutant and wild-type mRNAs in vivo

Experiment (plasmid) and mRNA	Relative activity[a] of:		
	C′	P	C
Expt 1 (pSC)[b]			
Wild type	154	100	10
C′ AUG	168	2	1
C′ GCG	3	219	106
Expt 2 (pGEM)[c]			
Wild type	122	100	22
C′ AUG	129 ± 7	4 ± 2	3 ± 1

[a] In both experiments, the amount of radioactivity in the P protein band of the wild-type construct was set at 100, and all other values are given relative to this.

[b] Values were determined by densitometry from the results in Fig. 3.

[c] The pGEM constructs were expressed by transfection in BHK cells superinfected with vTF7-3 (9). The proteins were labelled with [^{35}S]methionine and immunoprecipitated as described for Fig. 3. Radioactivity was quantitated by a Phosphorimager. Each value for C′ AUG is the average and range of two determinations carried out in separate but parallel transfections.

On the basis of other instances where non-AUG start codons are used, there appear to be two rationales for this situation. In some cases, the upstream non-AUG start codon is used to create a functionally distinct form of the protein. For example, in murine leukemia virus, the N-terminally elongated form of the *gag* protein (initiated by CUG) contains a signal sequence which targets it to the cell surface in an uncleaved form, whereas the AUG-initiated *gag* protein remains in the cytoplasm and is processed (22). Alternatively, the upstream non-AUG might be used to downregulate expression from downstream AUGs (12); for PIV1 and SEN, this would include the essential P protein (4). Here, a functionally distinct form of the protein would not be relevant. At present, we have no information on whether C′ and C have distinct subcellular localizations, as there is no antiserum that can distinguish them available as yet. However, we have recently found that the SEN C protein(s) strongly inhibits viral transcription (unpublished data) but not replication (4). It is therefore now possible to determine whether C′ and C have distinct properties, at least in this respect. The fact that the C′ protein is expressed far more abundantly than C for PIV1, whereas the opposite is true for SEN, can be reconciled more easily with the notion that they represent distinct but common forms of the same function.

We thank Catherine Stouder for excellent technical assistance, Laurent Roux (Geneva, Switzerland) and Hiroshi Komada (Mie, Japan) for antibodies, and Roberto Cattaneo and Martin Billeter (Zurich, Switzerland) for pSC.

This work was supported by a grant from the Swiss National Science Fund.

REFERENCES

1. **Acland, P., M. Dixon, G. Peters, and C. Dickson.** 1990. Subcellular fate of the Int-2 oncoprotein is determined by choice of initiation codon. Nature (London) **343**:662–665.
2. **Becerra, S. P., J. A. Rose, M. Hardy, B. M. Baroudy, and C. W. Anderson.** 1985. Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. Proc. Natl. Acad. Sci. USA **82**:7919–7923.
3. **Bernards, A., and S. M. de la Monte.** 1990. The ltk receptor tyrosine kinase is expressed in pre-B lymphocytes and cerebral neurons and uses a non-AUG translational initiator. EMBO J. **9**:2279–2287.
4. **Curran, J., R. Boeck, and D. Kolakofsky.** 1991. The Sendai virus P gene expresses both an essential protein and an inhibitor of RNA synthesis by shuffling modules via mRNA editing. EMBO J. **10**:3079–3085.
5. **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.
6. **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.
7. **Etkind, P. R., R. K. Cross, R. A. Lamb, D. C. Merz, and P. W. Choppin.** 1980. In vitro synthesis of structural and nonstructural proteins of Sendai and SV5 viruses. Virology **100**:22–33.
8. **Florkiewicz, R. Z., and A. Sommer.** 1989. Human basic fibroblast growth factor gene encodes four polypeptides: three initiate translation from non-AUG codons. Proc. Natl. Acad. Sci. USA **86**:3978–3981.
9. **Fuerst, T. R., E. G. Niles, F. W. Studier, and B. Moss.** 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. USA **83**:8122–8126.
10. **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.
11. **Gupta, K. C., and S. Patwardhan.** 1988. ACG, the initiator codon for a Sendai virus protein. J. Biol. Chem. **263**:8553–8556.
12. **Hann, S. R., M. W. King, D. L. Bentley, C. W. Anderson, and R. N. Eisenman.** 1988. A non-AUG translational initiation in c-myc Exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell **52**:185–195.

12a. **Huber, M., R. Cattaneo, P. Spielhofer, C. Örvell, E. Norrby, M. Messerli, J.-C. Perriard, and M. Billeter.** 1991. Measles virus phosphoprotein retains the nucleocapsid protein in the cytoplasm. Virology **185**:299–308.

13. **Kadowaki, H., T. Kadowaki, F. E. Wondisfond, and S. I. Taylor.** 1989. Use of polymerase chain reaction catalyzed by Taq DNA polymerase for site-specific mutagenesis. Gene **76**:161–166.
14. **Kingsbury, D. W.** 1991. The paramyxoviruses. Plenum Press, New York.
15. **Kozak, M.** 1989. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol. Cell. Biol. **9**:5073–5080.

15a. **Kozak, M.** 1990. Downstream secondary structure facilitates recognition of initiator codons by eucaryotic ribosomes. Proc. Natl. Acad. Sci. USA **87**:8301–8305.

16. **Kuroya, M., and N. Ishida.** 1953. Newborn virus pneumonitis (type Sendai). II. Report: the isolation of a new virus possessing hemagglutinin activity. Yokohama Med. Bull. **4**:217–233.
17. **Lyn, D., D. A. Gill, R. A. Scroggs, and A. Portner.** 1991. The nucleoproteins of human parainfluenza virus type 1 and Sendai virus share amino acid sequences and antigenic and structural determinants. J. Gen. Virol. **72**:983–987.
18. **Matsuoka, Y., J. Curran, T. Pelet, D. Kolakofsky, R. Ray, and R. W. Compans.** 1991. The P gene of human parainfluenza virus type 1 encodes P and C proteins but not a cysteine-rich V protein. J. Virol. **65**:3406–3410.
19. **Matsuoka, Y., and R. Ray.** 1991. Sequence analysis and expression of the human parainfluenza type 1 virus nucleoprotein gene. Virology **181**:403–407.
20. **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.

20a. **Mehdi, H., E. Ono, and K. C. Gupta.** 1990. Initiation of translation at CUG, GUG, and ACG codons in mammalian cells. Gene **91**:173–178.

21. **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.
22. **Prats, A.-C., G. de Billy, P. Wang, and J.-L. Darlix.** 1989. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. J. Mol. Biol. **205**:363–372.
23. **Prats, H., M. Kaghad, A.-C. Prats, M. Klagsbrun, J. M. Lélias, P. Liauzun, P. Chalon, J. P. Tauber, F. Amalric, J. A. Smith, and D. Caput.** 1989. High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons. Proc. Natl. Acad. Sci. USA **86**:1836–1840.
24. **Saris, C. J. M., J. Domen, and A. Berns.** 1991. The pim-1 oncogene encodes two related protein-serine/threonine kinases by alternative initiation at AUG and CUG. EMBO J. **10**:655–664.
25. **Sugihara, H., V. Andrisani, and P. M. Salvaterra.** 1990. Drosophila choline acetyltransferase uses a non-AUG initiation codon and full length RNA is inefficiently translated. J. Biol. Chem. **265**:21714–21719.
26. **Taira, M., T. Iizasa, H. Shimada, J. Kudoh, N. Shimizu, and M. Tatibana.** 1990. A human testis-specific mRNA for phosphoribosylpyrophosphate synthetase that initiates from a non-AUG codon. J. Biol. Chem. **265**:16491–16497.
27. **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.