Lack of Correlation between Sendai Virus P/C mRNA Structure and Its Utilization of Two AUG Start Sites from Alternate Reading Frames: Implications for Viral Bicistronic mRNAs†

Kailash C. Gupta,*,‡ Etsuro Ono,‡,§ and Xiulong Xu‡,||

Department of Immunology/Microbiology, Rush Medical College, 1653 West Congress Parkway, Chicago, Illinois 60612, and Department of General Surgery, Rush-Presbyterian St. Luke's Medical Center, Chicago, Illinois 60612

Received August 30, 1995; Revised Manuscript Received November 28, 1995⊗

ABSTRACT: The polycistronic P/C mRNA of Sendai virus encodes five proteins (C′, P, C, Y1, and Y2) each of which initiates from a distinct start site. Two major proteins, P and C, are expressed in approximately equimolar amounts from two consecutive AUGs in overlapping reading frames. To better understand the mechanism of expression of the C protein from a downstream AUG, site-directed mutants of the P/C mRNA were created and expressed in COS1 cells. The secondary structure of the mRNA was examined to determine whether the mRNA structure played any role in the synthesis of the C protein. Our results ruled out any significant involvement of the 5′ UTR, sequence contexts, secondary structure, distance between the start sites, and sequences downstream to the C-AUG. However, they are consistent with the concept that the synthesis of the C protein is primarily dependent on the orientation of its reading frame, i.e., +1 in relation to the upstream P reading frame. The downstream reading frame was translated poorly when it occurred in +2 orientation in relation to the upstream reading frame. Interestingly, all the known functional bicistronic mRNAs with overlapping reading frames from cytoplasmic RNA viruses have their downstream reading frame in +1 orientation relative to the upstream frame. We propose that the evolutionary conservation of the downstream reading frame in +1 orientation in these bicistronic mRNAs is important for its efficient translation.

The mechanism by which a eukaryotic ribosome identifies the start site for protein synthesis on an mRNA differs significantly from that used by a prokaryotic ribosome (Moldave, 1985). Recent studies have shown that the eukaryotic ribosome can identify an appropriate start site by alternate means. In the majority of mRNAs, translation initiation factor(s) (Sonenberg, 1988) or the 40S subunit of the eukaryotic ribosome (Kozak, 1991) binds to the 5′-capped end of an mRNA and slides toward the 3′ end in search of an initiation codon. Several lines of evidence have suggested that the efficiency with which an AUG codon is used is dependent on its sequence context. Sequence comparisons and mutational analyses have suggested the "optimum" start site for the eukaryotic ribosome as 5′ PuNNAUGPu (Cavener, 1987; Kozak, 1991). In less than 10% of mRNAs, because of the suboptimum nature of the 5′-proximal initiator AUG, some ribosomes scan past that AUG and may initiate translation from a downstream AUG(s) (Kozak, 1991). This "leaky" scanning of the ribosome allows synthesis of two or more proteins from in-frame or out-of-frame AUGs. Alternatively, the ribosome can bind internally to an mRNA in a cap-independent manner to access the translation start site as has been demonstrated for picornaviruses (Pelletier & Sonenberg, 1988; Jang et al., 1988; McBratney et al., 1993) and the mammalian immunoglobulin heavy-chain binding protein mRNAs (Macejak & Sarnow, 1990).

Studies with viral bicistronic mRNAs that initiate translation from two consecutive AUGs in overlapping reading frames have provided some novel information on translation initiation in eukaryotes. However, the mechanism by which the downstream translation start site is recognized in these mRNAs remains enigmatic. Alkhatib et al. (1988) showed, for the P/C mRNA of measles virus, that the upstream P-AUG had no effect on the level of C protein synthesis that initiated from a downstream AUG. These authors suggested a scanning-independent mechanism for the synthesis of the measles virus C protein. Williams and Lamb (1989), using influenza B virus NB/NA mRNA, observed an enhancing role of 5′-untranslated region in the synthesis of NB and NA proteins. However, for the s1 mRNA of reovirus, differential chain elongation rates (Fajardo & Shatkin, 1990) and sequences downstream of the start sites (Belli & Samuel, 1993) have been suggested to regulate the expression of the encoded proteins, s1 and p14. It is possible, therefore, that various structural attributes of a bicistronic mRNA as well as alternate translation mechanisms facilitate initiation from the downstream start site (Samuel, 1989).

The polycistronic P/C mRNA of Sendai virus (Figure 1), which is translated into at least five proteins in two overlapping reading frames during viral infection (Dillon & Gupta, 1989), offers a good model to determine the factors that affect translation initiation of the encoded proteins. Interestingly, the 5′-proximal translation start site of the C′ protein is an ACG triplet. The remaining four downstream

† This work was supported by grants from the American Cancer Society (NP817E) and the NIH (AI 30517).

* Corresponding author. Telephone: (312) 942−7768. Fax: (312) 226-6020.

‡ Rush Medical College.

§ Present address: Department of Veterinary Hygiene and Microbiology, Faculty of Veterinary Medicine, Hokkaido University, Sapporo, Japan 060.

|| Rush-Presbyterian St. Luke's Medical Center.

⊗ Abstract published in *Advance ACS Abstracts,* January 15, 1996.

0006-2960/96/0435-1223$12.00/0

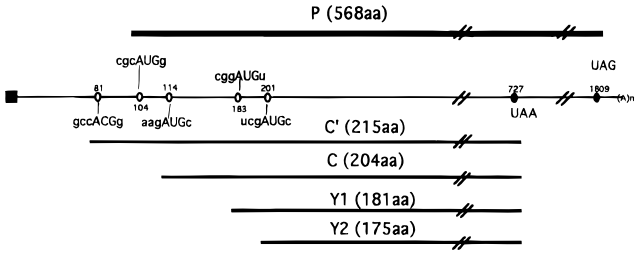

FIGURE 1: Schematic representation of the Sendai virus P/C mRNA and its coding frames. The open and closed circles, respectively, represent translation start and stop sites. The nucleotide position of each start codon, their sequence contexts, and the size of the encoded protein are given.

initiation codons are AUGs, each in a suboptimal sequence context (Curran & Kolakofsky, 1988; Gupta & Patwardhan, 1988) (Figure 1). This implies that they are "weak" start sites and are probably accessed by the "leaky" ribosomes (Kozak, 1988). Our previous studies have confirmed that the different start sites in the P/C mRNA are accessed by the leaky ribosomes (Gupta, 1987; Mehdi et al., 1990). However, the relative amounts of the various proteins synthesized from this mRNA can not be explained by the current tenets of the leaky ribosomal scanning model (Kozak, 1991). In the present paper, we show that the sequence contexts and other structural attributes of an mRNA which putatively affect translation initiation in monocistronic mRNAs do not significantly affect translation initiation from the two major start sites in the P/C mRNA. Importantly, our results suggest that the +1 orientation of the downstream C-reading frame may be important in its efficient translation.

MATERIALS AND METHODS

Construction of Recombinant Plasmids. Most of our recombinant DNA manipulations were performed as described by Sambrook et al. (1989). Recombinants pG1f1PC and pBC/PC contained the entire P/C gene of Sendai virus Z strain (Shioda et al., 1983). pG1f1PC and its mutants were used for site-directed mutagenesis and for the in vitro synthesis of P/C mRNA specific transcripts (Patwardhan & Gupta, 1988). pBC/PC and its mutants were used for transfection and expression in COS1 cells (Mehdi et al., 1990). During transfection, the P/C mRNA is expressed from a CMV early promoter in pBC/PC which adds 75 nucleotides to the 5′ terminus of the P/C mRNA (Cullen, 1986). Consequently, the wild-type P/C and all its mutant mRNAs contain heterologous 75 nucleotides at their 5′ termini on expression in COS1 cells.

P/C gene subcloning was primarily accomplished by using unique *Hind*III and *Bam*HI sites in pG1f1PC and pBC/PC. To facilitate subcloning of some constructs, we changed the original unique *Hind*III site to *Nhe*I in both recombinants by linker ligation. Site-directed mutants of pG1f1PC (Table 1) were generated by a previously described procedure (Patwardhan & Gupta, 1988). To insert synthetic oligonucleotide duplexes, a *Nco*I site was created at nucleotide 81 (C′-ACG to AUG, mutant P22) or at P-AUG (mutant P31/66) and a *Hind*III site at nucleotide 135 (mutant P21). Before creating a *Hind*III site at nucleotide 135, the original unique *Hind*III site at nucleotide 1 was changed to a *Nhe*I site (pBC/PC/*Nhe*) as discussed above (Table 2). Deletions from the 5′ end of the P/C mRNA to C′-AUG (PΔ22+) or to P-AUG (PΔ31) were obtained by digestion of mutants P22 or P31/66, respectively, with *Nco*I and *Hind*III. The small fragment was deleted; the ends of the large fragment were filled by Klenow polymerase and then ligated to a *Hind*III linker. The C′-AUG of PΔ22+ was mutagenized by site-directed mutagenesis to ACG to obtain PΔ22.

Duplex replacement mutants (D2, D3, D6, and D7) were created by inserting synthetic DNA duplexes between the *Nco*I (nucleotide 81) and *Hind*III (nucleotide 135) sites (Table 2). To reduce the distance between C′- and P-AUGs by three codons, mutant D2 was created by inserting a 44 bp oligonucleotide. Mutant D7 is identical to D2 except that five codons were deleted between C′ and P start sites by inserting a 38 bp oligonucleotide duplex. Similarly, mutant D3 was designed to reduce the distance between P- and C-AUGs by one codon. This mutant was created by replacing the *Nco*I/*Hind*III fragment with a 27 bp duplex molecule. D3 lacks 23 nucleotides between C′ and P start sites and 12 nucleotides downstream of the C start site. Mutant D6 was created by inserting a 21 bp duplex. It contains novel sequences between P and C start sites, but maintains the original distance. Like D3, it lacks 23 nucleotides between C′ and P start sites, but lacks 21 nucleotides downstream of the C start site. To avoid any context effects, the translation start sites are in optimum sequence context in all the duplex replacement mutants (**Pu**NN**AUGPu**).

The identity of the subcloned P/C gene, duplex insertions, and site-directed mutants was confirmed by sequencing the recombinants using P/C gene specific oligonucleotide primers and the Sequenase 2.0 kit (USB).

In Vitro Transcription. Transcripts of wild-type pG1f1PC or its mutants were synthesized in vitro using SP6 RNA polymerase (Melton et al., 1984). Briefly, 0.5 μg of template DNA linearized with *Bam*HI was transcribed in a reaction containing 50 mM Tris-HCl (pH 8.0), 4 mM $MgCl_2$, 2 mM spermidine, 1 mM each of ATP, CTP, GTP, and UTP, 10 units of RNasin (Promega), 10 mM DTT, and 20 units of SP6 RNA polymerase (USB) in a final volume of 25 μL. The mixture was incubated at 38 °C for 90 min, and a 2 μL aliquot of the reaction was analyzed in a 1% agarose gel containing 0.5 μg of ethidium bromide. The gel was visualized by UV for the integrity and quantity of transcripts. The transcripts were finally extracted with phenol/chloroform/isoamyl alcohol (25:24:1), precipitated with ethanol, dissolved in water, and quantitated by UV spectrophotometry.

Nuclease Digestion and Primer Extension Reaction. To determine single- and double-stranded regions in the wild-type and mutant transcripts, 0.5 μg of transcripts and 5 μg of yeast tRNA were equilibrated for 5 min at room temperature in RNase CV1 or nuclease S1 digestion buffers (Shelness & Williams, 1985). CV1 digestion buffer contained 120 mM KCl, 50 mM Tris-HCl (pH 8.3), 10 mM $MgCl_2$, and 1 mM DTT; and nuclease S1 buffer contained 100 mM NaCl, 25 mM NaOAc, and 1 mM $ZnCl_2$, pH 4.7. The amount of nucleases (CV1 and S1) necessary to obtain appropriate partial digestion of the transcripts was determined experimentally. After digestion at 25 °C for 10 min, 5 μg of carrier yeast tRNA was added, and the RNA was recovered following phenol/chloroform extraction and ethanol precipitation. RNA was air dried, dissolved in 10 μL of CV1 buffer, and used as template for primer extension reaction.

Four oligonucleotide sequencing primers (19−20 mers), complementary within the 250-nucleotide 5′-terminal region of the P/C mRNA, and spaced at intervals of about 60 nucleotides, were 5′-end-labeled with [γ^{32}P]ATP (3000 Ci/mmol) and T4 polynucleotide kinase. Nuclease-digested transcripts were reverse-transcribed with 3 pmol of a labeled primer, 1 mM each of dNTPs, and 5 units of AMV reverse transcriptase (RTase) at 42 °C for 90 min. Reactions were stopped with a gel loading solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol FF), and aliquots were analyzed in a 6% polyacrylamide sequencing gel. To precisely determine the site of cleavage, a sequence ladder of the template DNA with the corresponding primer was prepared and analyzed in parallel with the primer extension products.

Secondary Structure Modeling of P/C mRNA. The RNAFOLD program, developed by Educational & Scientific Software based on the algorithm of Martinez (1984) and the energy parameters of Freier et al. (1986), was employed to determine potential stem−loops in the wild-type and mutant transcripts. Secondary structure models of the 5′-proximal region of the P/C mRNA were created by taking into account the experimental results and the computer-generated potential structures.

Transfection, Metabolic Labeling, and Immunoprecipitation. The transfection protocol has been described in detail previously (Mehdi et al., 1990). Briefly, COS1 cells plated in 60 mm dishes were transfected following the DEAE-dextran method (Selden, 1990). Monolayers were washed with phosphate-buffered saline (PBS) and then transferred to 2 mL of DMEM containing 10% Nu-serum. Four micrograms of plasmid DNA in 20 μL of Tris-buffered saline (TBS) was mixed with 40 μL of 10 mg/ml DEAE-dextran (in TBS) and added to the medium. After incubation for 4 h, the medium was aspirated, and the cells were shocked by 10% DMSO in PBS for 2 min at room temperature, washed twice with PBS, and incubated for an additional 48 h in DMEM + 10% FBS.

Forty-eight hours after transfection or 24 h after Sendai virus infection (5 pfu/cell), cells were transferred to methionine-free DMEM for 1 h and then labeled for 2 h in the presence of 100 μCi/mL ^{35}S-Trans-label (ICN). Cells were washed with PBS and lysed in 1 mL of RIPA (radioimmunoprecipitation assay) buffer containing 20 mM Tris-HCl (pH 7.5), 2 mM EDTA, 0.15 M NaCl, 1% sodium deoxycholate, 1% Triton X-100, 0.1% SDS, 1 mM PMSF, 1 μg/mL leupeptin, and 1 μg/mL aprotinin. Nuclei and cellular debris were removed by centrifugation at 24000*g* for 30 min. Three hundred microliters of the lysate from the transfected cells or 30 μL from the infected cells was used for immunoprecipitation with 1 μL of the rabbit polyclonal antiserum raised against the carboxyl-terminal peptide of the C protein (Portner et al., 1985) and/or 2.0 μL of (1:10 dilution) monoclonal antibody against the P protein (Deshpande & Portner, 1985). Under these conditions of immunoprecipitation, the antibodies were in excess to antigens, and the levels of immunoprecipitated P and C proteins were in the linear range. Immunocomplexes were adsorbed to 15 μL of Pansorbin (Calbiochem). The bacterial pellet was washed 3 times with RIPA buffer, 500 μL each, suspended in 15 μL of 2 × Laemmli loading buffer, boiled for 3 min, and resolved in a 12.5% SDS−polyacrylamide gel (Laemmli, 1971). Gels were fixed, treated with Enhancer (NEN) for fluorography, dried, and exposed to X-ray films. Autoradiographs were scanned and data computed using a laser densitometer from Molecular Dynamics. The relative levels of the P and C proteins during Sendai virus infection and transient expression were highly comparable and reproducible. All the transfections have been repeated at least 3 times with similar results. In addition, the relative level of P and/or C protein in various mutants served as good controls for each others expression. Various mutations in the P/C mRNA had no effect on the levels of the P/C mRNA expression as analyzed by Northern blot hybridization (data not presented).

Table 1: Reading Frames in the P/C mRNA and Its Point Mutants[a]

		% Expression			
		C′	ΔC′	P	C
WT:	CAGCC **ACG** GCU UCG GCU ACA CUU ACC GC**A UGG** AUC AAG **AUG** CCU UCA (81, 104, 114; C′, P, C)	100	0	100	100
30:	CAGCC **ACG** GCU UCG GCU ACA CUU ACC GC**A UGG** AUC AAG **AUG** ⁂CU UCA	0	0	100	100
31:	CAGCC **ACG** GCU UCG GCU ACA CUU AC⁂ GC**A UGG** AUC AAG **AUG** CCU UCA	25	0	100	50
30/31:	CAGCC **ACG** GCU UCG GCU ACA CUU AC⁂ GC**A UGG** AUC AAG **AUG** ⁂CU UCA	0	0	100	50
31/70:	CAGCC **ACG** GCU ⁂⁂G GCU ACA CUU AC⁂ GC**A UGG** AUC AAG **AUG** CCU UCA	100	100	<5	<5
31/66:	CAGCC **ACG** GCU UCG GCU ACA CUU AC⁂ ⁂**A UGG** AUC AAG **AUG** CCU UCA	25	0	100	50
22:	CAGCC **A⁂G** GCU UCG GCU ACA CUU ACC GC**A UGG** AUC AAG **AUG** CCU UCA	1000	0	<10	<10
22/31:	CAGCC **A⁂G** GCU UCG GCU ACA CUU AC⁂ GC**A UGG** AUC AAG **AUG** CCU UCA	1000	0	<10	<10

[a] Translation start site for each protein is presented with boldface and underlined letters, and sites of mutations are represented with outline letters. The level of expression of each protein from the wild-type (WT) construct is assigned a value of 100%, and the level of expression of corresponding proteins from mutants is based on the WT values. As ΔC′ is not present in the WT construct, the value of this protein in mutant 31/70 is assigned as 100%.

RESULTS

Weak Sequence Context of the P-AUG Is Not Important for C Protein Synthesis. We and others have shown that the synthesis of the P and C proteins is about equimolar in infected as well as in transfected cells (Dillon & Gupta, 1989; Mehdi et al., 1990; Curran & Kolakofsky, 1988). The P-AUG (CGCAUGG) is 11 nucleotides upstream from the C-AUG (AAGAUGC), and both AUGs are putatively in "weak" sequence contexts (Kozak, 1991). To test whether the C start site is recognized by leaky ribosomes, we optimized the sequence context of the P-AUG by substituting the C residue at the −3 position with an A (P31). Similarly, the C-AUG context was optimized by substituting the C at the +4 position to a G (P30). Finally, both the P-AUG and C-AUG contexts were optimized in a single construct (P30/31) (Table 1). To assess the effects of these mutations on the synthesis of P and C proteins, the mutants were transiently expressed in COS1 cells. To minimize background bands, P and C proteins were immunoprecipitated independently from equivalent amounts of the same lysate (Dillon & Gupta, 1989). Although the strong sequence context of the P-AUG (P31) did not have any effect on the synthesis of the P protein, it decreased C protein synthesis by about 50% (Figure 2). Optimizing the C-AUG context (P30) had no significant effect on the synthesis of the C and P proteins, and the relative levels of these proteins were comparable to the wild type (PC). As expected from the above results, when both P- and C-AUG contexts were optimized (P30/31), the amount of P and C proteins expressed was similar to those observed with P31. Importantly, these results show that even when the upstream

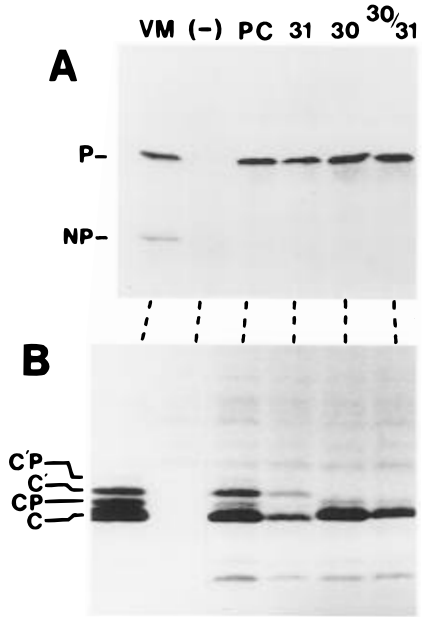

FIGURE 2: Expression of sequence context mutants in COS1 cells. P and C proteins were immunoprecipitated independently to reduce background. In mutants P31 and P30, P-AUG and C-AUG were, respectively, transferred to good context. Both AUGs were transferred to good context in mutant P30/31. Lane VM shows the immunoprecipitation from virus-infected cells while lane (−) shows immunoprecipitation from the mock-transfected cells. C′P and CP are the phosphorylated forms of C′ and C, respectively (Hendricks et al., 1993).

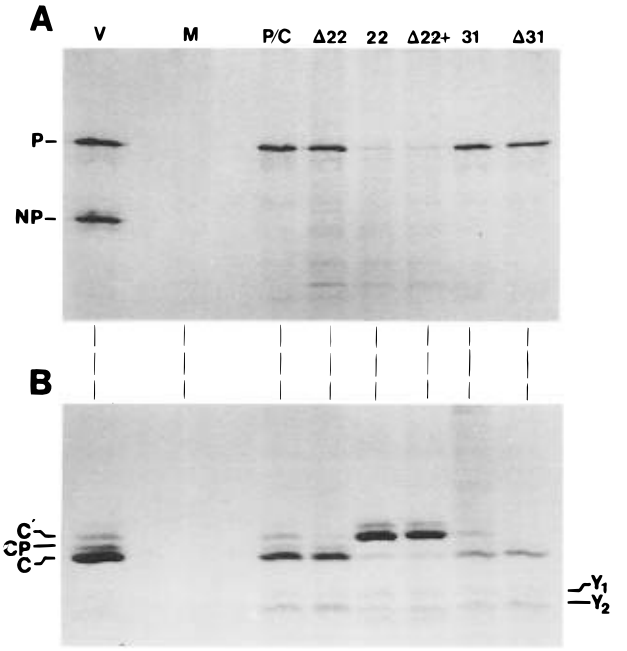

FIGURE 3: Expression of 5′ UTR deletion mutants of the P/C mRNA in COS1 cells. PΔ22 and PΔ31 lack 79 and 102 nucleotides of the 5′ UTR, respectively. Mutants PΔ22 and PΔ22+ carry ACG and AUG, respectively, at the C′ start site. Mock (M), virus-infected (V), and wild-type (P/C) lanes have been shown for comparison.

P-AUG was in optimal sequence context, a substantial amount of the C protein was synthesized, but no increase in the level of P synthesis occurred. Unexpectedly, P30 abrogated the translation from the upstream C′-ACG, and P31 reduced C′ protein synthesis to 25%. The basis of these results is not clear. However, our recent results suggested that it is most likely due to the conformational changes in the mRNA (Gupta et al., 1994).

We have previously demonstrated that P and C start sites are identified by ribosomes scanning from the 5′ end of the P/C mRNA (Gupta, 1987; Mehdi et al., 1990). To reconfirm that P and C start sites are recognized by ribosomal scanning, we expressed mutant P22 transiently. In P22, the C′-ACG start site has been changed to AUG, thereby creating an AUG codon in a strong sequence context upstream of P and C start sites. As expected, in mutant P22, synthesis of the C′ protein increased 10-fold while that of P and C proteins was reduced by 10-fold in comparison to the wild type (P/C) (Figure 3). The minor persisting amounts of P and C proteins are most likely due to the excess ribosomes scanning the P/C mRNA as we have previously proposed (Mehdi et al., 1990). These results further demonstrated that P and C start sites are recognized in a scanning-dependent manner. However, the presence of a substantial amount of the C protein in mutants with a strong upstream P-AUG (P31 and P30/31) suggested that some other factor(s) must be involved in the expression of C protein from the downstream start site.

The 5′ Untranslated Region (UTR) Plays No Role in the Synthesis of C Protein. To determine whether the 5′ UTR of the P/C mRNA regulates the synthesis of C protein, we constructed two mutants which lacked different lengths of the UTR. Mutant PΔ22+ lacks the 5′-terminal 79 nucleotides and has an AUG start codon for the C′ protein. This mutant expressed mainly the C′ protein and minor amounts of P and C as observed in the mutant P22 (Figure 3). The C′-AUG of PΔ22+ was changed back to the wild-type C′-ACG to obtain PΔ22. Transfection of this mutant yielded levels of P and C which were identical to the wild-type expression levels except that the level of C′ diminished to about 25% (Figure 3). It is interesting to note that a similar decrease in C′ synthesis was observed in mutant P31. We have no explanation for these unexpected results, but we have shown recently that initiation from a non-AUG codon is mRNA-specific (Gupta et al., 1994). Although these results ruled out any role of the 5′-proximal 79 nucleotides in regulating the synthesis of C protein, the remaining 23 nucleotides of the P/C mRNA upstream from the P-AUG could still play a role. To test this possibility, we deleted 102 nucleotides upstream from the P-AUG. This deletion mutant (PΔ31) was derived from the mutant P31 wherein the GCATGG was changed to CCATGG (P31/66) to create an *Nco*I site (Table 1). Creation of the *Nco*I site did not change the expression profile of P31 (data not shown). The *Hind*III/*Nco*I fragment was then deleted as described under Materials and Methods to generate PΔ31. Expression of PΔ31 was similar to that obtained with the mutant P31, except, as expected, the C′ protein was missing (Figure 3). These results clearly demonstrate that the 5′ UTR of the P/C mRNA is not involved in the expression of the C protein.

All our constructs are expressed from the early CMV promoter−enhancer region which carries a transcription start site 75 nucleotides upstream from the *Hind*III cloning site (Cullen, 1986). Therefore, all the P/C gene specific transcripts expressed in COS1 cells have a common 75-nucleotide region upstream of the P/C mRNA sequence. As this 75-nucleotide region is present in all the constructs including the deletion mutants and had no effect on the appropriate expression of the P/C gene, we believe that this extraneous region has no influence on the expression of P/C mRNA encoded proteins.

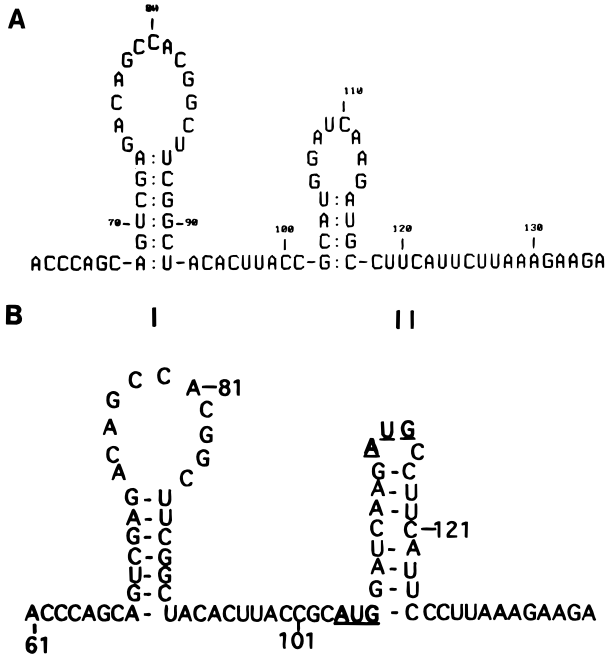

FIGURE 4: Hairpins in the 5′-proximal region of the P/C mRNA. Figure 4A shows the hairpins predicted by RNAFOLD (S&E Software), while Figure 4B represents the hairpins generated by experimental analysis of the mRNA with single- and double-strand-specific nucleases. P and C start sites have been underlined in Figure 4B.

5′-Proximal Region of the P/C mRNA Is Poorly Structured but Has a Potential for Modulation. To determine whether the secondary structure of the P/C mRNA plays a role in the expression of C protein, we experimentally analyzed the single- and double-stranded regions in the 5′-terminal 150 nucleotides (nucleotides 1−150) of the P/C mRNA which encompasses the start sites for C′, C, and P proteins. Single-strand-specific nuclease S1 and double-strand-specific nuclease CV1 were employed to partially digest P/C transcripts, and the digested transcripts were used as templates for reverse transcription using 5′-^{32}P-labeled antisense oligonucleotide primers (Shelness & Williams, 1985). As our initial experiments revealed that the 5′-proximal region of the P/C mRNA was highly sensitive to salt concentration, near-physiological salt conditions (120−160 mM KCl) were used for nuclease digestions to maintain the biological configuration of the RNA. With the help of RNAFOLD-generated potential stem−loops (Figure 4A) and our nuclease digestion results, two distinct hairpins were identified (Figure 4B). Hairpin I has a calculated thermal stability of −4.2 kcal/mol and encompasses the C′-ACG. Nucleotides 67−74 and 86−92, which together have the potential to form a base-paired region, were weakly susceptible to digestion with CV1 nuclease, whereas the intervening single-stranded region, nucleotides 75−84, was digested by nuclease S1 (Figure 5A). Hairpin II, which was not predicted by RNAFOLD (compare Figure 4A and Figure 4B), was identified around the P and C start codons with the nuclease digestion pattern. Figure 5B1 and Figure 5B2 show two continuous (106−113 and 118−125 nucleotides) partially digested areas by CV1. Between them is a region of four nucleotides not digested by S1. This hairpin has a calculated thermal stability of +1.2 kcal/mol. Thus, C-AUG is in the loop, and P-AUG is adjacent to the start of the stem. Spontaneous terminations of primer extensions occurred at nucleotides C103, A104, and C111. Sequence-specific terminations at CA and CU doublets were also observed by Shelness and Williams (1985).

The weak stabilities of the hairpins are consistent with the complete digestion of the 5′ UTR with nuclease S1 in the absence of salt (data not presented). It is interesting to note that while hairpin I was predicted to be stronger than hairpin II, the nuclease digestion patterns for hairpin II were more distinct. Although for most of the structural analysis we used in vitro synthesized uncapped transcripts, the structures in the P/C mRNA were confirmed using total cytoplasmic RNA from Sendai virus infected cells (data not presented). The primer extension products from the P/C mRNA revealed the presence of a 5′-terminal cap, as expected, by carrying an additional nucleotide at their 5′ termini (Gupta & Kingsbury, 1984). The hairpins identified in the 5′ proximal region of the P/C mRNA are rather weak and, therefore, have a potential to be modulated.

To directly test the possibility that hairpin II is involved in the expression of C protein from a downstream AUG, we used mutants in which hairpin II was disrupted. Mutants D3 and D6, originally created to test the role of distance and sequences between P and C start sites, and the role of C-AUG downstream sequences, had hairpin II completely disrupted by deletion of 12 (in D3) or 21 (in D6) nucleotides downstream of C-AUG (Table 2). The nuclease digestion patterns of transcripts of these mutants confirmed the loss of hairpin II. No alternate secondary structures were detected in this region (data not presented). As described earlier, the −2 G residue of the P-AUG context was changed to C to create an *Nco*I site for facile insertion of duplexes between *Nco*I and *Hind*III ends. This alteration did not affect the relative amounts of P and C proteins (data not presented). Expression of P and C proteins from D3 (Figure 6) was comparable to that obtained with the mutant P30/31. These results also indicated that the distances between P and C start sites and 12 nucleotides downstream of the C-AUG were not important. However, in mutant D6 the level of C protein was about 30% that of the P protein. As described under Materials and Methods, D6 has novel sequences between P- and C-AUGs and has a 21-nucleotide deletion downstream of C-AUG. Despite several described changes in the D3 and D6 mRNAs, the C protein was expressed rather efficiently. These results suggest that the nucleotides downstream of the C start site, sequence between the two start sites, and the hairpin II may have a modulatory role but not the determining role in C protein synthesis.

Plus One Reading Frame May Be Important for C Protein Synthesis. The above results ruled out the possibilities that the start site sequence contexts, the 5′ UTR, the sequences immediately downstream of the start sites, sequences between the start site, or the secondary structure plays a dominant role in the expression of C from the P/C mRNA. In addition, the mutants (D2, D3, D6, D7, and several not presented) showed that the distance between the start sites was not important. However, an overview and analysis of all the results presented in this paper and our related unpublished results suggested that the architecture of the P and C reading frames in the P/C mRNA plays a role in the expression of the C protein. All mutants that had their second reading frame in the +1 orientation were translated efficiently from both frames while the mutants that had their second reading frame in the +2 orientation were translated primarily from

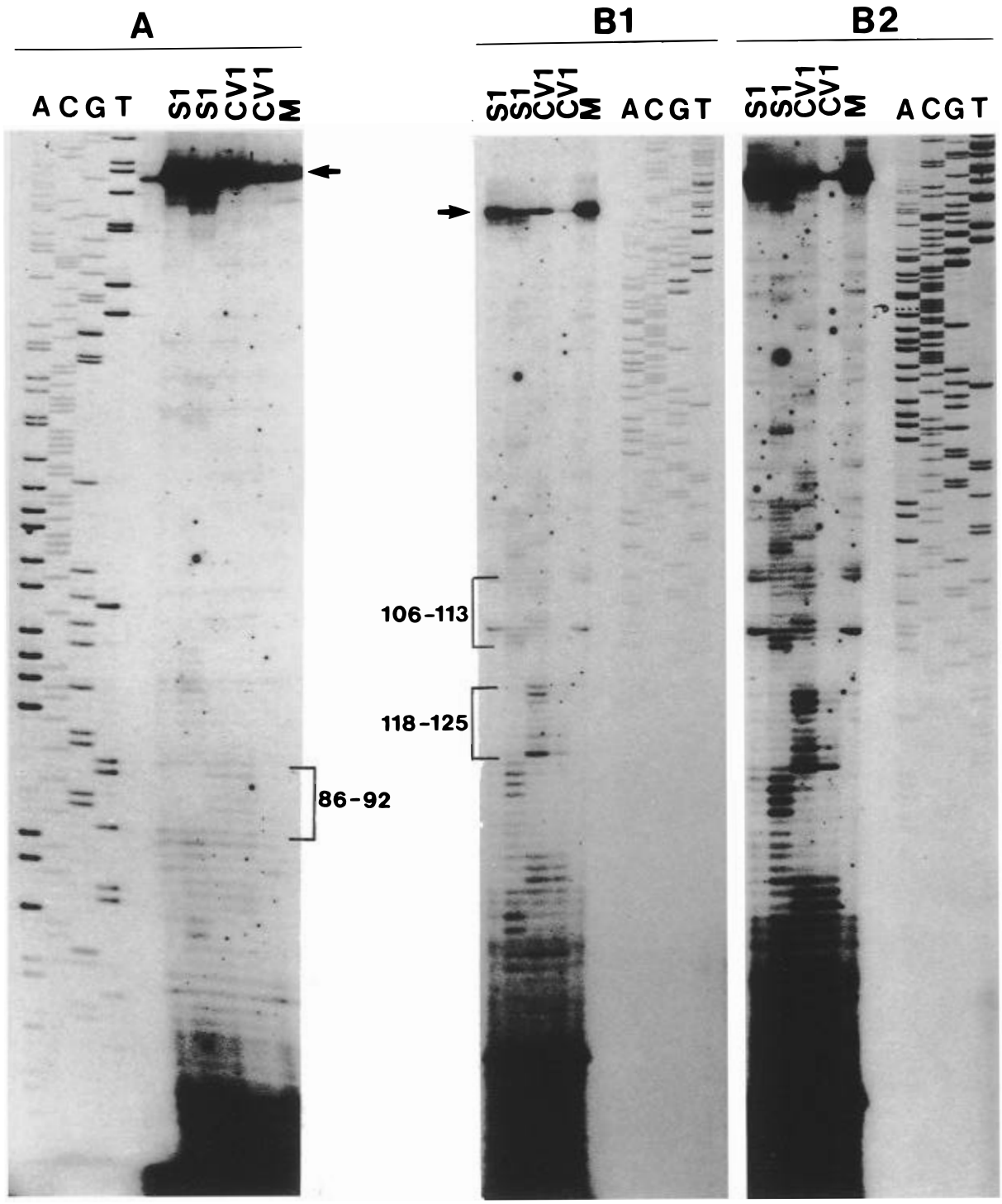

FIGURE 5: Analysis of nuclease S1- and CV1-sensitive sites in the 5′-proximal region of the P/C transcripts. Wild-type P/C transcripts (0.5 μg) were partially digested with 0.1 (right lanes) or 0.01 (left lanes) unit of nuclease S1 or nuclease CV1 at room temperature for 10 min. The primer extension products templated by the nuclease digestion products were resolved by 6% polyacrylamide sequencing gels. Arrows indicate the 5′ end of the transcripts. CV1-sensitive sites have been shown by the position of nucleotides in the P/C mRNA. Figure 5A shows the CV1 site proximal to the 5′ end of the mRNA. Figures 5B1 and 5B2 are shorter and longer exposures of the same gel. Primer extension products from the mock-treated transcripts (lanes M) were run alongside the S1 and CV1 lanes. The sequence ladders (run in parallel) were prepared with the same primer as used for primer extension reactions. Primer P47 (5′- TCCGAGAGCGACTCTCGTCC) complementary to the P/C mRNA from nucleotides 164−183 was used for panel A, and primer P60 (5′-GTCCCCTCCTATGTCAGTT) complementary from nucleotides 223−241 was used for panel B.

the upstream reading frame. In particular, for mutants P22 and P31/70, in which the P reading frame is in a +2 position relative to the C′ frame, little P synthesis was detected (Figure 7). Similarly, in mutant 31/70, where the P-AUG is in the +2 orientation to a novel ΔC′-AUG, very little P protein was detected (Figure 7). The ΔC′-AUG is two codons downstream of the C′-ACG. In contrast, mutants P31, P31/30, D3, and D6 (see Figures 2 and 6), in which the C reading frame is in the +1 position relative to the P frame, expressed significant amounts of the C protein.

Due to the design of the P/C mRNA, it is not possible to transfer downstream C-AUG to the +2 frame in relation to the P frame without encountering several stop codons in the P reading frame. Therefore, to test indirectly the concept

Table 2: Duplex Replacement Mutants of the P/C mRNA[a]

		% Expression		
		C′	P	C
WT:	CAGCC ACG GCU UCG GCU ACA CUU ACC GCA UGG AUC AAG AUG CCU UCA (C ←24→ P ←11→ C)	100	100	100
D2:	CAGCC AUG GCU CUU---Δ---ACA GCA UGG AUC AAG AUG CCU UCA (C ←15→ P ←11→ C)	1000	<5	<5
D7:	CAGCC AUG ---Δ---GCA GCA UGG AUC AAG AUG CCU UCA (C ←9→ P ←11→ C)	500	<5	<5
D3:	CAGCC ---Δ---A UGG GC--Δ--G AUG CCU UCA (P ←8→ C)	0	100	50
D6:	CAGCC ---Δ---A UGG CAA AAC AUG CCU UCA (P ←11 novel→ C)	0	100	30

[a] Translation start sites are presented by boldface and underlined letters and point mutations by outline fonts. Nucleotide distances between the start sites of the wild type (WT) and various duplex replacement mutants are shown. Deletion of nucleotides is presented by dashed and Δ symbols. The level of expression of each WT protein is assigned as 100%, and the level of expression of corresponding proteins from mutants is based on the WT values.

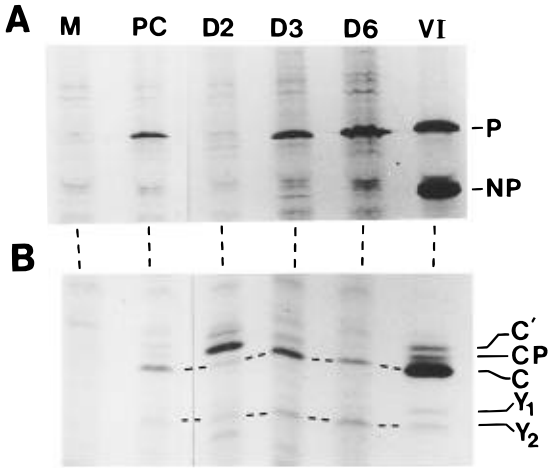

FIGURE 6: Expression of duplex replacement mutants created to disrupt the stem-loop structures around the P and C start sites. Details of the mutants are presented in Table 2. ^{35}S-Labeled P (Figure 6A) and C (Figure 6B) proteins were immunoprecipitated independently from Sendai virus infected (VI), mock (M), or wild-type (PC) lysates. Proteins were resolved in a 12% SDS−discontinuous polyacrylamide gel. In mutant D2, P-AUG is in the +2 orientation in relation to C′-AUG. In mutants D3 and D6, C-AUG is in the +1 orientation relative to P-AUG.

that the position of the downstream reading frame is primarily important in its synthesis, we transferred the C reading frame upstream of the P which in turn puts the P frame in the +2 position, as in the point mutants P31/70 and P22/31. Two duplex replacement mutants, D2 and D7, were constructed in which the distance between C′-AUG and P-AUG was reduced from the original 24 to 15 and 9 nucleotides, respectively (Table 2). These mutants expressed only C′ with no detectable expression of P (Figure 8). On the contrary, reducing the distance between P- and C-AUGs by one codon (mutant D3) and deleting both upstream and downstream sequences (mutants D3 and D6) but keeping the C frame in the +1 orientation allowed significant amounts of C protein synthesis (see Figure 6 and Table 2). For a better comparison and to avoid any AUG-context-related affect, we used mutants with their AUGs in good sequence context. All our results support the concept that for the efficient expression

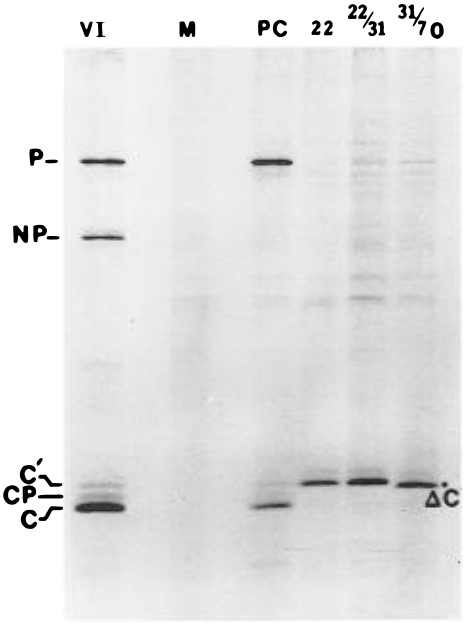

FIGURE 7: Expression of mutants (P22, P22/31, and P31/70) wherein the P-AUG is in the +2 orientation relative to C′-AUG or ΔC′-AUG. Mutant sequences are provided in Table 1. Both P and C proteins were immunoprecipitated together. In mutants P22/31 and P31/70, starting AUGs of both reading frames are in good sequence context. Immunoprecipitations from mock-transfected (M) and virus-infected (VI) cell lysates are shown for comparison.

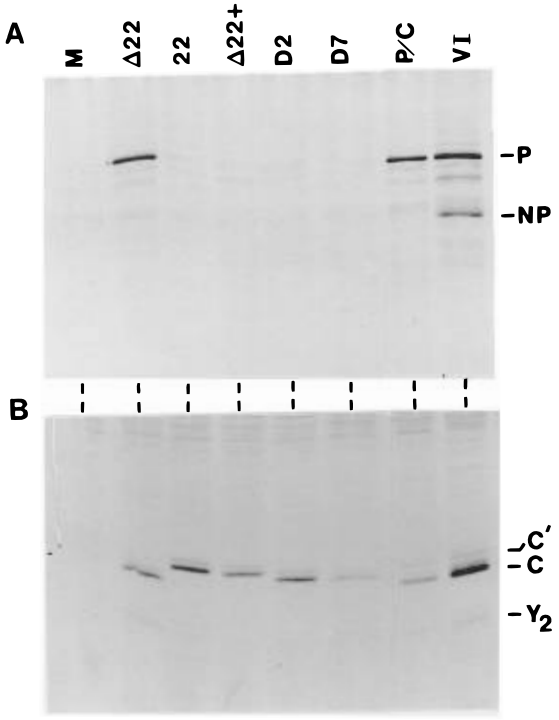

FIGURE 8: Expression of duplex replacement mutants (D2 and D7) created to examine P protein synthesis from the +2 downstream reading frame. Other mutants (P22 and PΔ22+) where the P reading frame is in the +2 orientation were also analyzed. While D2 and D7 both have AUGs in good sequence context, the context of P-AUG in P22 and PΔ22+ is in the original "weak" context. For comparison, expression of mutant P22 is shown where the C reading frame is in the +1 orientation as in the wild type (P/C). Immunoprecipitations from mock-transfected (M) and virus-infected (VI) cell lysates are also shown.

of the C protein from P/C mRNA, the downstream open reading frame must be in the +1 orientation to the upstream one.

DISCUSSION

Efficient Translation from the +1 Reading Frame. The mechanism of translation initiation from alternate start sites on a bicistronic mRNA is poorly understood. Previous studies (Gupta & Kingsbury, 1985; Alkhatib et al., 1988; Williams & Lamb, 1989; Curran & Kolakofsky, 1989; Fajardo & Shatkin, 1990; Doohan & Samuel, 1993) have shown that translation of bicistronic mRNAs with overlapping reading frames is rather unusual and various elements of mRNAs may be involved in regulating the initiation of two proteins from the same mRNA species. In an effort to understand the mechanism which regulates initiation from the C start site in the Sendai virus P/C mRNA, we created several mutants and expressed these mutants in COS1 cells. Our results showed that AUG sequence contexts, sequences upstream and downstream of AUGs, the distance between the start sites, and the local secondary structure do not play any major role in the synthesis of C protein from the downstream AUG. Curiously, mutants wherein several sequence alterations were incorporated (D3 and D6) expressed C protein efficiently. The only common characteristic among the mutants that expressed the downstream reading frame efficiently was the presence of a second AUG in a +1 orientation. However, the design of the reading frames in the P/C mRNA does not allow us to test this concept directly, as transferring the C-AUG to the +2 frame would terminate the P-reading frame after four codons. Moreover, several stop codons follow this stop codon. Due to this inherent problem, we tested this concept indirectly by transferring the P reading frame to the +2 orientation with respect to the C′ reading frame. In all the constructs (P22/31, P31/70, D2, and D7) where the P reading frame is in the +2 orientation to the upstream C′-AUG reading frame, P protein was not synthesized or was synthesized minimally.

Significantly, an examination of sequences of all other viral bicistronic mRNAs that initiate from consecutive AUGs (paramyxoviruses as well as other cytoplasmic RNA viruses in GENBANK) revealed that the +1 position of the downstream AUG is maintained. For example, bicistronic mRNAs from influenza virus B (Shaw et al., 1983), bunyamwera virus (Bishop et al., 1982), bovine coronavirus (Senanayake et al., 1992), reovirus (Cashdollar, et al, 1985; Ernst & Shatkin, 1985), simian rotavirus (Mitchell & Both, 1988), vesicular stomatitis virus (Hudson et al., 1986; Spiropoulou & Nichols, 1993), chicken infectious bursal disease virus (Spies, 1989), measles virus (Bellini et al., 1985), human parainfluenza virus 3 (Spriggs & Collins, 1986), rinderpest virus (Baron et al., 1993), canine distemper virus (Barrett et al., 1985), and pneumonia virus of mice (Barr et al., 1994) all have +1 orientation for their downstream reading frame. Consistent with our results, the sequences around their start sites are not conserved, and they do not reveal any conserved predictive secondary structures. Moreover, the distances between the two start sites are highly variable, from a minimum of 8 nucleotides in the NB/NA mRNA of influenza B virus (Shaw et al., 1983) to 121 nucleotides in the P/C mRNA of pneumonia virus of mice (Barr et al., 1994). Taken together, the evolutionary conservation of the reading frame architecture in all viral bicistronic mRNAs and our results lead us to propose that the +1 reading frame of the C protein in Sendai virus P/C mRNA is important for its efficient synthesis.

Possible Access of the +1 Frame by Frame-Shifting. One possible mechanism for efficient recognition of the +1 AUG could be by frame-shifting. Frame-shifting during translation has been observed both in prokaryotes and in eukaryotes. Efficient (30%) +1 frame-shift in *E. coli* was observed to autoregulate the synthesis of ribosome release factor RF-2 (Craigen & Caskey, 1987). In contrast, the +2 frame-shift occurred with an efficiency of only about 2% (Falahee et al., 1988). Plus-one frame-shifting was suggested to express low levels of full-length human thymidine kinase (TK) polypeptide from its single-base frame-shift mutant (Hwang et al., 1994). Frame-shifting (−1) is utilized by retroviruses at the gag-pol region (Jacks et al., 1987) and by the coronavirus avian infectious bronchitis virus at the F1−F2 junction (Brierley et al., 1987) to express regulated amounts of the respective viral polymerases. Similarly, −1 frame-shifting has been described for *Saccharomyces cerevisiae* double-strand RNA virus (ScV) (Tzeng et al., 1992). However, the +1 frame-shift in bicistronic viral mRNAs would differ from the above examples in that the frame-shift in bicistronic mRNAs would have to occur at an AUG codon rather than any sense codon. The results presented in this paper suggest that the +1 shift is more efficient than the +2 shift. Conceivably, frame-shifting at an AUG codon is more efficient than any other sense codon. We hypothesize that +1 reading frame-shift is conserved in viruses as it is more efficient as compared to the +2 reading frame-shift. Presumably, a shift of one nucleotide by the ribosome is easier than a shift of two.

In retroviruses and coronavirus, a downstream pseudoknot and a heptameric shifty sequence are required for frame-shifting (Brierley et al., 1989; Chamorro et al., 1992). In contrast, the minimal requirement, for +1 frame-shift in yeast retrotransposon, Ty, to synthesize TYA−TYB fusion, is only seven nucleotides, suggesting involvement of two specific tRNAs in this event (Belcourt & Farabaugh, 1990). However, site-specific mutations and gross alterations of the Sendai virus P/C mRNA as well as the comparative sequence analysis of the bicistronic viral mRNAs have not revealed any sequence requirements except that the AUGs of the overlapping reading frames are consecutive. We hypothesize, therefore, that the +1 frame in the bicistronic mRNAs is accessed by ribosomes which are committed (like imprinting) to the upstream reading frame, but these ribosomes can frame-shift efficiently when they encounter an AUG in the +1 frame. Two alternate scenarios are possible; in one, the ribosome would initiate translation at the upstream AUG and frame-shift with a certain frequency at the downstream AUG, transferring the initiated peptide at the amino terminus of the downstream protein. Alternatively, the ribosome may commit to the first AUG without actually initiating protein synthesis, and when it encounters the downstream AUG, it switches the frame. Future investigations, particularly the amino-terminal sequences of the downstream protein and construction of synthetic bicistronic mRNAs, would elicit the mechanism of +1 frame choice translation in the bicistronic mRNAs.

In conclusion, contrary to expectations, we have shown that structural features of the Sendai virus P/C mRNA are not involved in regulating the expression of the two major proteins (P and C) from consecutive AUGs. Instead, our work revealed that most likely +1 orientation of the downstream AUG in relation to the first AUG is important.

Although this concept is not easily testable, evolutionary conservation of this translation strategy in all the known bicistronic viral mRNAs supports this concept.

ACKNOWLEDGMENT

We thank Dr. Allen Portner for his generous supply of P monoclonal antibodies.

REFERENCES

Alkhatib, G., Massie, B., & Briedis, D. J. (1988) *J. Virol. 62*, 4059−4069.
Baron, M. D., Shaila, M. S., & Barrett, T. (1993) *J. Gen. Virol. 74*, 299−304.
Barr, J., Chambers, P., Harriott, P., Pringle, C. R., & Easton, A. J. (1994) *J. Virol. 68*, 5330−5334.
Barrett, T., Shrimpton, S. B., & Russell, S. E. (1985) *Virus Res. 3*, 367−372.
Belcourt, M. F., & Farabaugh, P. J. (1990) *Cell 62*, 339−352.
Belli, B. A., & Samuel, C. E. (1993. *Virology 193*, 16−27.
Bellini, W. J., Englund, G., Rozenblatt, S., Arnheiter, H., & Richardson, C. D. (1985) *J. Virol. 53*, 908−919.
Bishop, D. H. L., Gould, K. G., Akashi, H., & Clerx-van Haaster, C. M. (1982) *Nucleic Acids Res. 10*, 3703−3713.
Brierley, I., Bournsnell, M. E. G., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. K., & Inglis, S. C. (1987) *EMBO J. 6*, 3779−3785.
Brierley, I., Digard, P., & Inglis, S. C. (1989) *Cell 57*, 537−547.
Cashdollar, L. W., Chmelo, R. A., Wiener, J. R., & Joklik, W. K. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 24−28.
Cavener, D. (1987) *Nucleic Acids Res. 15*, 1353−1361.
Chamorro, M., Parkin, N., & Varmus, H. E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 713−717.
Craigen, W. J., & Caskey, C. T. (1987) *Biochimie 69*, 1031−1041.
Cullen, B. R. (1986) *Cell 46*, 973−982.
Curran, J. A., & Kolakosky, D. (1988) *EMBO J. 7*, 245−251.
Curran, J. A., & Kolakofsky, D. (1989) *EMBO J. 8*, 521−526.
Deshpande, K. L., & Portner, A. (1985) *Virology 140*, 125−134.
Dillon, P., & Gupta, K. C. (1989) *J. Virol. 63*, 974−977.
Doohan, J. P., & Samuel, C. E. (1993) *J. Biol. Chem. 268*, 18313−18320.
Ernst, H., & Shatkin, A. J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 48−52.
Fajardo, E. J., & Shatkin, A. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 328−332.
Falahee, M. B., Weiss, R. B., O′Connor, M., Doonan, S., Gesteland, R. F., & Atkins, J. F. (1988) *J. Biol. Chem. 263*, 18099−18103.
Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, Caruthers, M. H., Neilson, T., & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373−9377.
Gupta, K. C. (1987) *J. Biol. Chem. 262*, 7492−7496.
Gupta, K. C., & Kingsbury, D. W. (1984) *Nucleic Acids Res. 12*, 3829−3841.
Gupta, K. C., & Kingsbury, D. W. (1985) *Biochem. Biophys. Res. Commun. 131*, 91−97.
Gupta, K. C., & Patwardhan, S. (1988) *J. Biol. Chem. 263*, 8553−8556.
Gupta, K. C., Ono, E., Ariztia, E. V., & Inaba, M. (1994) *Biochem. Biophys. Res. Commun. 201*, 567−573.
Hendricks, D. D., Ono, E., Seyer, J. M., & Gupta, K. C. (1993) *Virology 197*, 471−474.
Hudson, L. D., Condra, C., & Lazzarini, R. A. (1986) *J. Gen. Virol., 67*, 1571−1579.
Hwang, C. B. C., Horsburgh, B., Felosi, E., Roberts, D. S., Digard, P., & Coen, D. M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5461−5465.
Jacks, T., Townsley, K., Varmus, H. E., & Majors, J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4298−4302.
Jang, S. K., Davies, M. V., Kaufman, R. J., & Wimmer, E. (1989) *J. Virol. 63*, 1651−1660.
Kozak, M. (1988) *Cell 47*, 481−483.
Kozak, M. (1991) *J. Cell Biol. 115*, 887−903.
Laemmli, U. K. (1970) *Nature 227*, 650−685.
Macejak, D. G., & Sarnow, P. (1990) *Enzyme 44*, 310−319.
Martinez, H. M. (1984) *Nucleic Acids Res. 12*, 323−334.
McBratney, S., Chen, C. Y., & Sarnow, P. (1993) *Curr. Opin. Cell Biol. 5*, 961−965.
Mehdi, H., Ono, E., & Gupta, K. C. (1990) *Gene 91*, 173−178.
Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., & Green, M. R. (1984) *Nucleic Acids Res. 12*, 7035−7056.
Mitchell, D. B., & Both, G. W. (1988) *Nucleic Acids Res. 16*, 6244.
Moldave, K. (1985) *Annu. Rev. Biochem. 54*, 1109−1149.
Patwardhan, S., & Gupta, K. C. (1988) *J. Biol. Chem. 263*, 4907−4913.
Pelletier, J., & Sonenberg, N. (1988) *Nature (London) 334*, 320−325.
Portner, A., Gupta, K. C., Seyer, J. M., Beachy, E. M., & Kingsbury, D. W. (1986) *Virus Res. 6*, 109−121.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Samuel, C. E. (1989) Prog. Nucl. Acid Res. *Mol. Biol. 37*, 127−153.
Selden, R. F. (1987) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brents, P., Kingston, R. E., Moore, D. M., Seidman, J. G., Smith, J. A. & Struhl, K., eds.) pp 941−943, Wiley, New York.
Senanayake, S. D., Hofmann, M. A., Maki, J. L., & Brian, D. A. (1992) *J. Virol. 66*, 5277−5283.
Shaw, M. W., Choppin, P. W., & Lamb, R. A. (1983) *Proc. Natl. Acad. Sci. U.S..A. 80*, 4879−4883.
Shelness, G. S., & Williams, D. L. (1985) *J. Biol. Chem. 260*, 8637−8646.
Shioda, T., Hidaka, Y., Kondo, T., Shibuta, H., Nomoto, A., & Iwasaki, K. (1983) *Nucleic Acids Res. 11*, 7317−7330.
Sonenberg, N. (1988) *Prog. Nucleic Acid Res. Mol. Biol. 35*, 173−207.
Spies, U. (1989) *Nucleic Acids Res. 17*, 7982−7992.
Spiropoulou, C. F., & Nichol, S. T. (1993) *J. Virol. 67*, 3103−3110.
Spriggs, M. K., & Collins, P. L. (1986) *J. Gen. Virol. 67*, 2705−2719.
Tzeng, T.-H., Tu, C.-L., & Bruenn, J. A. (1992) *J. Virol. 66*, 999−1066.
Williams, M. K., & Lamb, R. A. (1989) *J. Virol. 63*, 28−35.

BI9520646