Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 263, No. 18, Issue of June 25, pp. 8553-8556, 1988
© 1988 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.

ACG, the Initiator Codon for a Sendai Virus Protein*

(Received for publication, December 22, 1987)

Kailash C. Gupta‡ and Satish Patwardhan

From the Department of Immunology/Microbiology, Rush-Presbyterian-St. Luke's Medical Center, Chicago, Illinois 60612

Deletion and site-directed point mutants of the polycistronic P/C mRNA of Sendai virus revealed that one of the nonstructural proteins of this virus, the C′ protein, initiates from an ACG codon. This ACG codon occurs in an optimum sequence context and precedes the first AUG of the P/C mRNA. The results presented in this communication are consistent with the concept that the ribosomes scan the P/C mRNA to initiate several proteins from its different initiator codons. The arrangement of several weak initiator codons in tandem in an mRNA, *i.e.* non-AUG in optimum sequence context and AUG in suboptimum sequence context, may represent an alternate means to regulate gene expression in eukaryotes and their viruses.

Several non-AUG initiator codons are recognized by prokaryotic ribosomes (Gold *et al.*, 1981; Kozak, 1983). In contrast, no distinct examples for the utilization of non-AUG initiator codons in eukaryotes have been identified (Kozak, 1983). Yeast ribosomes have been shown to recognize UUG and AUA triplets as the initiator codons in mRNAs from artificial gene constructs (Zitomer *et al.*, 1984). A proposition for the possible utilization of an ACG initiator codon in higher eukaryotes was advanced on the basis of the nucleic acid sequence and protein products of adeno-associated virus (Becerra *et al.*, 1985). Bacteriophage T7 gene 0.3 in which the original AUG codon was substituted with ACG codon was observed to initiate protein synthesis in a wheat germ lysate translation system (Anderson and Buzash-Pollert, 1985). Similarly, based on cell-free translation and *in vivo* expression of mouse dihydrofolate reductase mRNA with AUG replaced with ACG, Peabody (1987) suggested that an ACG triplet may be recognized as an initiator codon in higher eukaryotes. In this communication, we show that an ACG triplet of a viral mRNA is recognized as an initiator codon by the eukaryotic ribosomes for the synthesis of a virus-specific protein.

We have been trying to understand initiation of translation of three proteins from a single viral mRNA species (Gupta and Kingsbury, 1985; Gupta, 1987). The mRNA is encoded by the P/C gene of Sendai virus, a cytoplasmic negative-strand RNA virus. The P/C mRNA in Sendai virus-infected cells (Lamb and Choppin, 1978) and in cell-free translation systems (Etkind et al., 1980) is translated into three proteins, P, C, and C′, from overlapping reading frames (Giorgi *et al.*, 1983; Shioda *et al.*, 1983). The P protein is a phosphoprotein which is putatively involved in the RNA-dependent RNA polymerase activity associated with the virus. The C and C′ proteins are nonstructural basic proteins (pI ~ 9.0) whose functions are not known. Two additional proteins, Y_1 and Y_2, have been detected as *in vitro* translation products of the P/C mRNA (Curran *et al.*, 1986; Patwardhan and Gupta, 1988) and in Sendai virus-infected cells.[1] The P/C mRNA has four 5′-proximal AUGs which initiate the P, C, Y_1, and Y_2 proteins, respectively (Fig. 1) as determined by site-directed mutagenesis of each of these AUGs (Patwardhan and Gupta, 1988). However, the initiator codon for the C′ protein remained elusive. Peptide analyses (Etkind *et al.*, 1980), immunoprecipitation studies (Portner *et al.*, 1986), and site-directed mutagenesis of the 5′-proximal AUGs (Patwardhan and Gupta, 1988) have shown that the C′ and C proteins are translated from the same reading frame, albeit starting from different initiator codons. Our previous site-directed mutants of the putative AUG initiator codons of the P/C mRNA suggested that the C′ protein may initiate from a non-AUG initiator codon (Patwardhan and Gupta, 1988). Here we demonstrate that the non-AUG initiator codon for the C′ protein is an ACG triplet which occurs in an optimum sequence context upstream from the first AUG in the P/C mRNA (Fig. 1).

EXPERIMENTAL PROCEDURES

Subcloning of the Sendai virus P/C gene into a multipurpose plasmid vector, pG1f1+, has been described previously (Patwardhan and Gupta, 1988). Oligonucleotide-directed site-specific mutants were generated using uridine containing single-stranded DNA templates of the P/C clone (pG1f1PC) essentially following the protocol we described elsewhere (Patwardhan and Gupta, 1988). The mutagenic oligonucleotides used to generate the mutants are presented in Table I. The mutants were named after the mutagenic primers. The mutants were characterized by plasmid DNA sequencing (Chen and Seeburg, 1985) (data not presented) and by *in vitro* translation of the SP6 RNA polymerase-directed transcripts of the P/C gene in a rabbit reticulocyte lysate system. The translation products were further confirmed by immunoprecipitation of the C, C′, and cognate proteins from the translation reactions (Patwardhan and Gupta, 1988). Proteins were analyzed by electrophoresis in 12% denaturing polyacrylamide gels (Laemmli, 1970).

For the amino-terminal sequence analysis of the [^{35}S]methionine-labeled C′ protein, wild-type P/C gene transcripts were translated in a wheat germ lysate system in the presence of oxaloacetate and citrate synthetase to block amino-terminal acetylation of the translated proteins (Palmiter, 1977). The labeled proteins were resolved by electrophoresis in a 12% preparative denaturing mini-gel containing 0.1% SDS (Laemmli, 1970). Immediately after the electrophoresis, gel was blotted onto an activated glass fiber filter (Aebersold *et al.*, 1986) using 25 mM *N*-ethylmorpholine buffer, pH 8.3. The labeled proteins were detected by autoradiography, and the band corresponding to the C′ protein was excised and was directly subjected to sequential Edman degradation in an automatic peptide sequenator (Applied Biosystems, model 477A/120A).

RESULTS

Delimiting the Region for the Initiator Codon of the C′ Protein—We have shown that none of the four 5′-proximal AUGs initiate the C′ protein and have suggested that C′ initiates from a non-AUG initiator codon (Patwardhan and Gupta, 1988). Based on immunoprecipitation of only the C

* This work was supported by Grant DMB 8609238 from the National Science Foundation (to K. C. G.) and Grant MV 360 from the American Cancer Society (to K. C. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed.

[1] P. J. Dillon and K. C. Gupta, unpublished results.

TABLE I

Specifications of mutagenic oligonucleotides

Oligonucleotides are complementary to the P/C gene. Mismatched nucleotide is underlined. Numbering of the P/C gene nucleotides came from Shioda *et al.* (1983).

Name	Sequence 5′ → 3′	Position on P/C gene	Targeted nucleotide and purpose
P21	ACTTC AGAAG CTTCT TTAA	126–144	135 Creation of *Hin*dIII site
P22	GCCGA AGCCA TGGCT GTCT	73–91	82 Mutate ACG → ATG
P23	GCCGA AGCCC TGGCT GTCT	73–91	82 Mutate ACG → AGG
P25	TGGCT GTCTA GACTG CTGG	63–81	72 Mutate GAG → TAG
P26	AGTGT AGCCT AAGCC GTGG	79–97	88 Mutate TCG → TAG

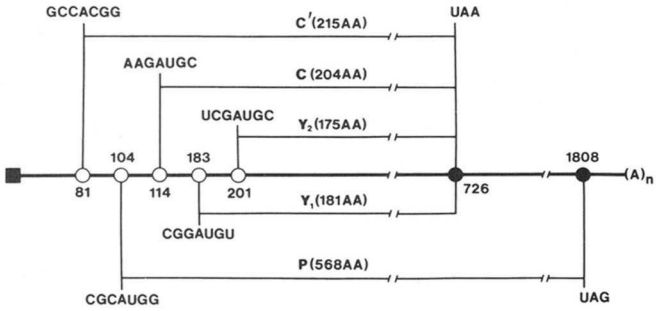

FIG. 1. **Graphic presentation of the P/C mRNA of Sendai virus.** The initiator codons utilized in the P/C mRNA expression have been assigned to the respective proteins. *Open circles* represent the initiator codons, and the *closed circles* represent the terminator codons. The *solid square* represents the 5′-capped terminus of the mRNA. Sequence context and position of each of the initiator codons are presented.

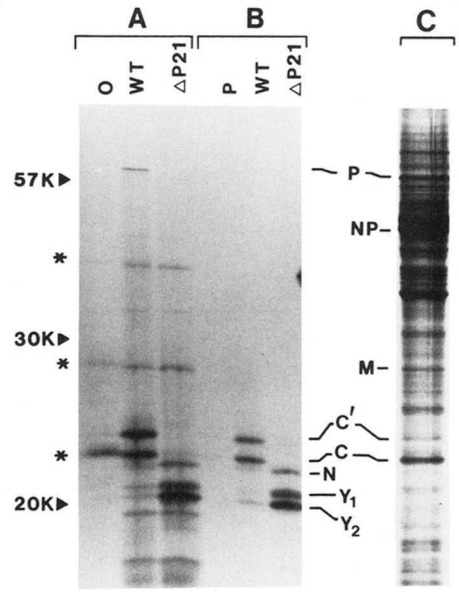

FIG. 2. **Electrophoretic analysis of *in vitro* translation products of capped transcripts derived from the wild-type (*WT*) and the deletion mutant ΔP21.** *A* represents the total translation products, and *B* represents the proteins precipitated with a rabbit antibody directed to the carboxyl terminus of the C proteins. *Lane O* represents the translation products without any input RNA, and *lane P* refers to the immunoprecipitated material from wild-type reaction with preimmune rabbit serum. *Asterisks* show the bands which appear in the absence of any input RNA (*A*). *C* presents metabolically ^{35}S-labeled proteins from Sendai virus-infected monkey kidney (LLC-MK2) cells. *P*, phosphoprotein; *NP*, nucleocapsid protein; *M*, matrix protein.

protein with an antibody specific to the amino terminus of the C protein and on the mobility of C and C′ proteins in denaturing gels, Curran *et al.* (1986) suggested that the C′ protein is smaller than the C protein. These observations are contrary to what has been observed and predicted by us (Gupta and Kingsbury, 1985; Gupta, 1987; Patwardhan and Gupta, 1988). To resolve this discrepancy, we decided to localize the region where the initiator codon for the C′ protein resides.

Accordingly, we created a mutant of the P/C gene in which the 5′-proximal 125 nucleotides were deleted. The deletion mutant (ΔP21) was generated by first creating a *Hin*dIII site at position 132 by substituting an A at position 135 with C (Table I). Following the creation of the *Hin*dIII site (mutant P21), a *Hin*dIII fragment (125 base pairs) was deleted, generating the mutant ΔP21 which lacked 5′-proximal 125 nucleotides of the P/C gene. Capped transcripts from the mutant ΔP21 were translated in a rabbit reticulocyte lysate system. The translation products showed that synthesis of the P, C, and C′ proteins was abolished in the deletion mutant (Fig. 2*A*). These results imply that the P, C, and C′ proteins initiate within the 5′-proximal 125 nucleotides. The disappearance of the P and C proteins is consistent with our recent observations (Patwardhan and Gupta, 1988) that the P and C proteins initiate from the AUGs at nucleotide positions 104 and 114, respectively, in the P/C mRNA (Fig. 1). These results further strengthen our hypothesis that the C′ protein initiates from a non-AUG initiator codon.

Though synthesis of the P, C, and C′ proteins in ΔP21 was abolished, three bands below the position of the C protein became prominent. To identify these proteins, we used antiserum raised against a peptide corresponding to the carboxyl terminus of the C proteins (Portner *et al.*, 1986) to specifically precipitate the proteins containing that sequence. The Y_1, Y_2, and a protein designated as N immunoprecipitated from the ΔP21 translation products (Fig. 2*B*). The precipitation of Y_1 and Y_2 was expected, as we have demonstrated that Y_1 and Y_2 initiate from AUGs at positions 183 and 201, respectively, and occur in the C reading frame. The increased synthesis of the Y_1 and Y_2 proteins in ΔP21 is consistent with the ribosome scanning hypothesis (Kozak, 1984) in that the deletion of the upstream initiator codons (for P, C, and C′) enhanced translation initiation from the two downstream AUGs. Although a faint band corresponding to the N protein is observed among wild-type translation products, its synthesis is particularly enhanced (~10-fold) in ΔP21 (Fig. 2, *A* and *B*). We believe that following the deletion of the first two AUGs (at 104 and 114), ribosomes recognize yet another non-AUG initiator

codon while scanning the ΔP21 mRNA before they reach AUG codons at 183 and 201. Indeed, the search for a non-AUG initiator codon between nucleotide positions 126 and 183 revealed that a CUG codon at position 138 in the C reading frame is in an optimum sequence context (AUUCUGA). Initiation of translation at this CTG triplet is consistent with the deduced size (197 amino acids) and electrophoretic mobility of the N protein (Fig. 2, *A* and *B*). However, it is important to note that the N protein has been detected so far only in *in vitro* translation products. Only further studies will determine whether this protein is produced *in vivo* and has any biological significance. Importantly, the above results delimited the initiation of the C′ protein with the 5′-proximal 125 nucleotides of the P/C mRNA.

Deciphering the Initiator Codon for the C′ Protein—The 5′-proximal 125 nucleotides of the P/C mRNA contain only two AUGs, one at the nucleotide position 104 and another at 114 (Shioda *et al.*, 1983; Giorgi *et al.*, 1983; Gupta and Kingsbury, 1984) (Fig. 1). We have already shown that the AUGs at 104 and 114 initiate the P and C proteins, respectively (Patwardhan and Gupta, 1988). Consequently, the search for a non-AUG initiator codon like those observed in prokaryotes (Gold and Stormo, 1987) within the 5′-proximal 125 nucleotides of the P/C mRNA revealed an ACG codon in a good sequence context (GCCACGG) at the nucleotide position 81 (Fig. 1). Furthermore, this codon lies in the same reading frame as the C protein and would provide an amino-terminal extension of 11 amino acids to the C protein, consistent with the electrophoretic mobility of the C′ protein.

To directly test if the ACG codon at position 81 is the initiator codon for the C′ protein, we generated point mutants: P22 and P23, (Table I). In mutants P22 and P23, we altered the ACG codon to AUG and AGG, respectively. It was expected that changing the ACG initiator codon to AGG (in P23) would abolish the synthesis of the C′ protein. On the other hand, altering ACG to AUG (in P22) will position an AUG in a good sequence context upstream of the other AUG codons of the P/C mRNA. Hence, this mutation might abolish the synthesis of the P, C, Y_1, and Y_2 proteins. The analysis of translation products and their immunoprecipitations are consistent with these predictions (Fig. 3, *A* and *B*). When ACG was changed to AUG (in P22), synthesis of the C protein was completely abolished; however, low level synthesis of the P protein persisted (Fig. 3*A*). It is possible that some of the scanning ribosomes did not stop at the new AUG codon and reached the downstream P-AUG.

To confirm that ACG is the initiator codon for the C′ protein, we generated mutants, P25 and P26, each having a terminator codon in the C reading frame (Table I). Mutant P25 carries a terminator codon (UAG) two codons upstream from the ACG triplet, while mutant P26 carries a terminator codon (UAG) two codons downstream from the ACG. We expected that P26 would abolish the synthesis of the C′ protein, while the P25 would not. The *in vitro* translation results confirm our predictions (Fig. 3, *A* and *B*).

Further, we tested the validity of our results by sequencing the [^{35}S]methionine-labeled C′ protein synthesized in a wheat germ lysate system. ^{35}S was released at cycles 1 and 12 (data not shown). These results confirmed that the C′ protein is indeed the C protein that carries 11 additional amino acids at its amino terminus. Finally, to determine unambiguously that the C′ protein studied by us is indeed identical to the *in vivo*-produced C′ protein, we analyzed the *in vitro* translation products and proteins from Sendai virus-infected cells by two-dimensional gel electrophoresis (Hsu and Kingsbury, 1982). Mobility of the *in vitro*-synthesized C′ protein (pI ~ 9) was

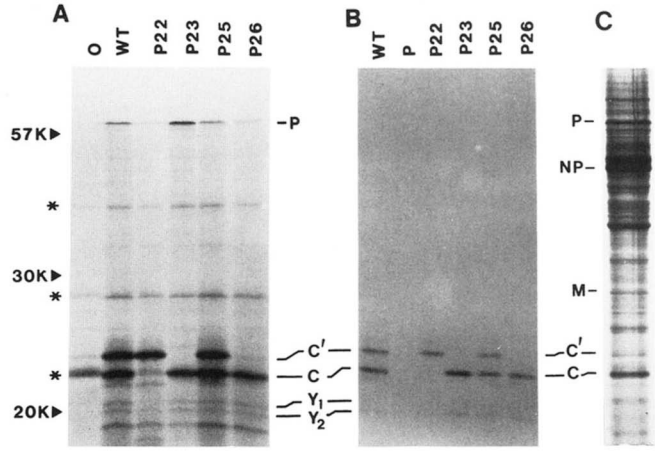

FIG. 3. **Electrophoretic analysis of *in vitro* translation products of wild type (*WT*) (pG1F1PC) and its point mutants, P22, P23, P25, and P26.** *A* represents the total translation products, and *B* represents the proteins precipitated with a rabbit antibody, as described in the legend to Fig. 2. *Lane O* represents the translation products without any input RNA, and *lane P* represents the immunoprecipitated material from wild-type reaction with preimmune rabbit serum. *Asterisks* show the bands which appear in the absence of any input RNA (*A*). *C* presents metabolically [^{35}S]methionine-labeled proteins from Sendai virus-infected monkey kidney cells. *P*, phosphoprotein; *NP*, nucleocapsid protein; *M*, matrix protein.

identical to the one produced in Sendai virus-infected cells (data not shown). Taken together, our results unambiguously demonstrate that ACG triplet at nucleotide position 81 is the initiator codon for the C′ protein.

DISCUSSION

The results presented in this communication demonstrate that ACG is the initiator codon for the C′ protein of Sendai virus. With the ACG triplet serving as the initiator codon, the C′ protein is 11 amino acids longer than the C protein (Fig. 1). As has been observed for the proteins initiating with the non-AUG initiator codons (Gold and Stormo, 1987; Peabody, 1987), the C′ protein initiates with a methionine. It is of interest to note that previously we had difficulty in explaining why an antisense oligonucleotide (20-mer) which spans the first two AUGs of the P/C mRNA abolished the synthesis of the P and C proteins but stimulated the synthesis of the C′ protein (Gupta and Kingsbury, 1985; Gupta, 1987). The discovery of the ACG initiator codon for the C′ protein upstream from the initiator codons of the P and C proteins helps to explain this observation. Since the ACG initiator codon is encountered by the scanning ribosomes prior to the AUG codons and because the downstream AUGs are obscured by the duplex formation in the presence of the oligonucleotide, the ribosomes preferentially initiate at the ACG codon. This explanation is consistent with our hypothesis that the scanning ribosomes have a "reach up" range of at least 20 nucleotides (Gupta, 1987). It is noteworthy that the ACG and the cryptic CUG (for the N protein) codons utilized in the P/C mRNA translation initiation occur in optimum sequence contexts. All of our previous and present observations are consistent with the basic tenets of the ribosome scanning model for translation initiation in eukaryotes (Kozak, 1984).

The ACG initiator codon of the P/C mRNA is utilized more efficiently *in vitro*, both in rabbit reticulocyte lysate (Gupta and Kingsbury, 1985) and in wheat germ extract,[2] than *in*

[2] K. C. Gupta and S. Patwardhan, unpublished results.

vivo (Curran *et al.*, 1986). Moreover, the utilization of the ACG initiator codon is enhanced by increasing the concentration of Mg^{2+} and/or spermidine during *in vitro* translation (Gupta and Kingsbury, 1985). Similar observations have been reported by Peabody (1987). The reason for this differential behavior of ribosomes is not clear; however, it is possible that Mg^{2+} and spermidine affect the fidelity of the initiation factors under *in vitro* conditions. Alternately, Mg^{2+} and spermidine may induce and stabilize conformation of an mRNA, allowing less efficient initiator codons to be recognized by ribosomes (Gupta and Kingsbury, 1985).

As discussed in the Introduction, known examples of the non-AUG initiator codons in eukaryotes are few and mostly derived from the artificial constructs of mRNAs. A capsid mRNA of adeno-associated virus and the P/C mRNA of Sendai virus are rare examples of naturally occurring mRNAs containing non-AUG initiator codons. These examples open up the possibility of unrecognized non-AUG initiator codons and, possibly, proteins in eukaryotic viruses and cells. It is important to note that the non-AUG codons are known to be recognized with rather low efficiency (Gold and Stormo, 1987), and it may be this reason that they are found upstream of the AUG initiator codons both in adeno-associated and Sendai viruses so that they can be recognized by the scanning ribosomes. It is conceivable that the proteins initiating from non-AUG initiator codons will be synthesized in small amounts and may represent regulatory proteins. The identification of more natural examples of non-AUG initiator codons will be necessary before their importance in eukaryotes can be generalized.

Acknowledgments—We thank Dr. Ka-Leung Ngai, Biotechnology Facilities, Northwestern University, for synthesis of oligonucleotides and amino-terminal sequencing of the C′ protein, Dr. Mark E. Peeples and Patrick J. Dillon for suggesting improvements in the manuscript, and Alice F. Robinson for manuscript preparation.

REFERENCES

Aebersold, R. H., Teplow, D. W., Hood, L. E., and Kent, S. B. H. (1986) *J. Biol. Chem.* **261,** 4229–4238
Anderson, C. W., and Buzash-Pollert, E. (1985) *Mol. Cell. Biol.* **5,** 3621–3624
Becerra, S. P., Rose, J. A., Hard, M., Baroudy, B. M., and Anderson, C. W. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 7919–7923
Chen, E. J., and Seeburg, P. H. (1985) *DNA* **4,** 165–170
Curran, J. A., Richardson, C., and Kolakofsky, D. (1986) *J. Virol.* **57,** 684–687
Etkind, P. R., and Cross, R. K., Lamb, R. A., Merz, D. C., and Choppin, P. W. (1980) *Virology* **100,** 22–33
Giorgi, C., Blumberg, B. M., and Kolakofsky, D. (1983) *Cell* **35,** 829–836
Gold, L., and Stormo, G. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhart, F. C., ed) pp. 1302–1307, American Society for Microbiology, Wash. D. C.
Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S., and Stormo, G. (1981) *Annu. Rev. Microbiol.* **35,** 365–403
Gupta, K. C. (1987) *J. Biol. Chem.* **262,** 7492–7496
Gupta, K. C., and Kingsbury, D. W. (1984) *Nucleic Acids Res.* **12,** 3829–3841
Gupta, K. C., and Kingsbury, D. W. (1985) *Biochem. Biophys. Res. Commun.* **131,** 91–97
Hsu, C. H., and Kingsbury, D. W. (1982) *Virology* **120,** 225–234
Kozak, M. (1983) *Microbiol. Rev.* **47,** 1–45
Kozak, M. (1984) *Nucleic Acids Res.* **12,** 3873–3893
Laemmli, U. K. (1970) *Nature* **227,** 680–685
Lamb, R. A., and Choppin, P. W. (1978) *Virology* **84,** 469–478
Palmiter, R. (1977) *J. Biol. Chem.* **252,** 6388–6393
Patwardhan, S., and Gupta, K. C. (1988) *J. Biol. Chem.* **263,** 4907–4913
Peabody, D. S. (1987) *J. Biol. Chem.* **262,** 11847–11851
Portner, A. P., Gupta, K. C., Seyer, J. M., Beachey, E. H., and Kingsbury, D. W. (1986) *Virus Res.* **6,** 109–121
Shioda, T., Hidaka, Y., Kando, T., Shibuta, H., Nomoto, A., and Iwasaki, K. (1983) *Nucleic Acids Res.* **11,** 7317–7330
Zitomer, R. S., Walthall, D. A., Raymond, B. C., and Hollenberg, C. P. (1984) *Mol. Cell. Biol.* **4,** 1191–1197