MOLECULAR AND CELLULAR BIOLOGY, Dec. 1985, p. 3621–3624
0270-7306/85/123621-04$02.00/0

Vol. 5, No. 12

Can ACG Serve as an Initiation Codon for Protein Synthesis in Eucaryotic Cells?

CARL W. ANDERSON* AND ELIZABETH BUZASH-POLLERT
Biology Department, Brookhaven National Laboratory, Upton, New York 11973

Received 15 July 1985/Accepted 10 September 1985

An ACG codon, which replaces the AUG codon used to initiate the synthesis of bacteriophage T7 gene 0.3 protein, was shown to function as a low-efficiency initiation codon in a wheat germ cell-free protein-synthesizing system.

In higher eucaryotic organisms, only AUG has been found to serve as an initiation codon for protein synthesis (11), although inefficient initiation in yeast cells at UUG and AUA was recently reported (18). In contrast, several codons can initiate protein synthesis in *Escherichia coli*; however, the utilization of non-AUG initiation codons is rare (reviewed in reference 11).

Adeno-associated virus (AAV) is a defective parvovirus that replicates in human tissue culture cells coinfected with adenovirus. A recent study of the three AAV capsid proteins suggested that the synthesis of one (AAV-B) initiates at an ACG codon (4; S. P. Becerra, J. A. Rose, M. Hardy, B. Baroudy, and C. W. Anderson, Proc. Natl. Acad. Sci. USA, in press). This hypothesis, which is based on the known sequence of the AAV2 genome (16), amino-terminal sequence analysis of the AAV-B capsid protein (4; Becerra et al., in press), and knowledge of the structure of capsid protein mRNAs (10, 12), explains the observed ratio of synthesis (ca. 1:20) of the AAV-B and AAV-C capsid proteins (5; Becerra et al., in press). AAV-B and AAV-C have the same carboxy terminus and share a sequence of 503 amino acids, but the B protein has 65 amino acids of additional sequence at its amino terminus. Immediately preceding the alanine codon specifying the amino terminus of the AAV-B protein is an ACG codon; no upstream AUG codons are found in known AAV-B capsid protein mRNAs (4; Becerra et al., in press).

If ACG were to serve (rarely) as an initiation codon for protein synthesis, in favorable circumstances, one might observe initiation at ACG in cell-free protein-synthesizing extracts produced from eucaryotic cells. We are unaware of any such reports, but if initiation at ACG were inefficient, e.g., 5% that of AUG at the equivalent position, in few instances would it have been detected. There is, however, one simple system in which in vitro initiation at an ACG codon might be detected.

Previously, we found that three bacteriophage T7 early mRNAs, although not capped by 7-methyl guanosine, were accurately translated in cell-free protein-synthesizing systems derived from rabbit reticulocytes or from wheat germ (1, 2). One of the T7 proteins synthesized efficiently in these systems is the product of gene 0.3. This 116-amino-acid protein inactivates the *E. coli* DNA restriction system; however, phage which do not produce active 0.3 gene protein can grow efficiently on restriction-deficient *E. coli* strains (17).

Two classes of mutants that inefficiently express 0.3 protein have been described (6). One class, represented by

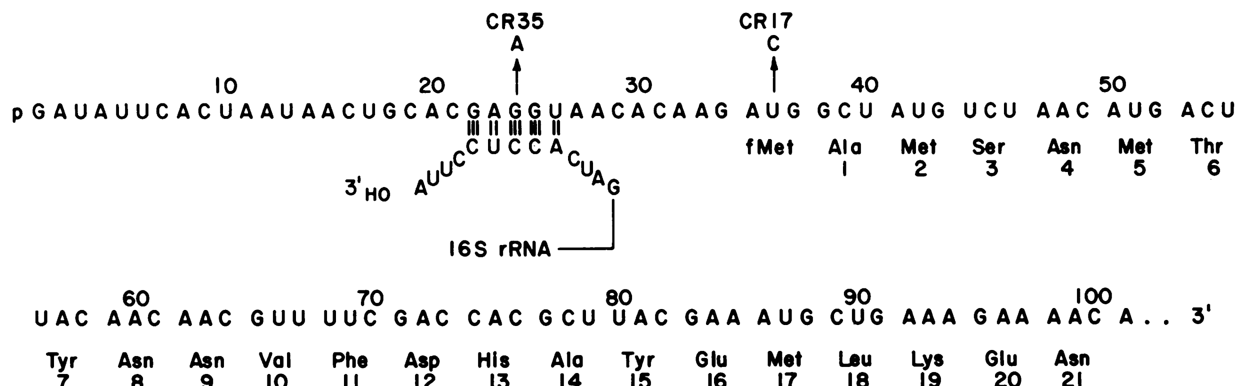

FIG. 1. Nucleotide and amino acid sequence near the 5' end of bacteriophage T7 0.3 protein mRNA. The 5' end of 0.3 protein mRNA is formed by RNase III-mediated cleavage of a larger polycistronic transcript. The nucleotide sequence 22 to 26 bases from the 5' end of 0.3 protein mRNA is complementary to a sequence near the 3' end of *E. coli* 16S rRNA; this complementary sequence is known as the Shine-Dalgarno region. The mutations found in phage CR17 and CR35 are indicated. In *E. coli*, protein synthesis initiates at the AUG codon at nucleotides 31 to 33 with formyl-methionyl tRNA; the initiating formyl-methionine is subsequently removed. From Dunn et al. (6), with permission.

* Corresponding author.

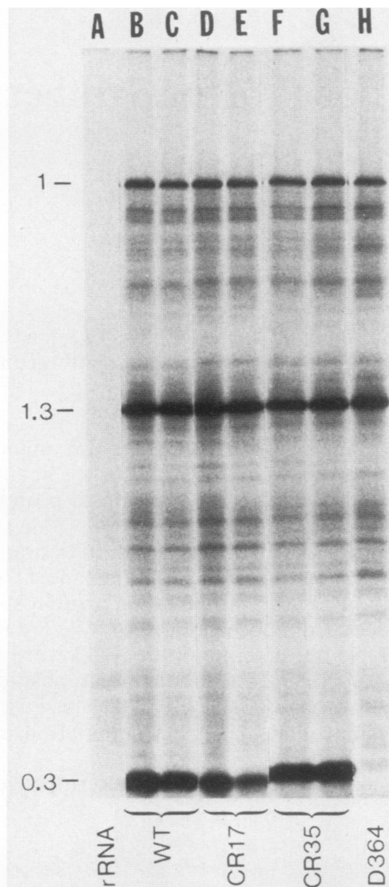

FIG. 2. Translation of bacteriophage T7 early mRNA in a wheat germ cell-free system. Autoradiogram of a sodium dodecyl sulfate-polyacrylamide gel showing the [^{35}S]methionine-labeled products synthesized in a wheat germ-derived cell-free protein-synthesizing system programmed with early T7 mRNA (made in vitro) from wild-type and bacteriophage T7 mutants CR17 (initiation codon for 0.3 protein altered to ACG) and CR35 (altered Shine-Dalgarno recognition sequence). The tracks shown are cell-free translations programmed with *E. coli* rRNA (A), T7 wild-type (WT) early mRNA at 50 μg/ml (B), T7 WT mRNA at 12 μg/ml (C), T7 mutant CR17 (ACG) mRNA at 50 μg/ml (D), T7 mutant CR17 mRNA at 12 μg/ml (E), T7 mutant CR35 mRNA (Shine-Dalgarno) at 50 μg/ml (F), T7 mutant CR35 mRNA at 12 μg/ml (G), and T7 deletion mutant D364 (makes no 0.3 protein) mRNA at 50 μg/ml (H). The concentration of processed T7 RNA required to saturate the protein synthetic capacity of the wheat germ cell-free system is greater than 100 μg/ml (2). The positions of the T7 gene 1 product (T7 RNA polymerase; molecular weight, 98,092), gene 1.3 product (T7 DNA ligase; molecular weight, 41,133), and gene 0.3 protein (molecular weight, 13,678) are marked at the left (9).

mutant CR35, is altered in the Shine-Dalgarno ribosome recognition region; the other, represented by CR17, has the 0.3 gene AUG initiation codon altered to ACG. The location of these mutants with respect to the sequence of the 5′ end of 0.3 gene mRNA is shown in Fig. 1. In vivo, phage carrying either mutation produce from 1/6th to 1/12th the normal amount of gene 0.3 protein; a similar reduction was observed in vitro by using an *E. coli* cell-free protein-synthesizing system (6).

We examined the products obtained in a wheat germ cell-free protein-synthesizing system programmed with mRNA from mutant CR17. Unequivocal evidence that protein synthesis can initiate at an ACG codon was obtained.

We prepared early T7 mRNAs by transcribing in vitro T7 DNA with *E. coli* RNA polymerase and processing the transcribed precursor to individual mRNAs with *E. coli* RNase III (2, 8). This RNA was used to program the wheat germ (3) or rabbit reticulocyte cell-free protein-synthesizing system. Both systems produce the gene 0.3 and 1.3 proteins in approximately the same relative proportions when programmed with wild-type mRNA, CR17 mRNA, or mRNA from CR35 (Fig. 2). Quantitative analysis of the results shown in Fig. 2 indicated that CR17 mRNA yielded approximately twofold less gene 0.3 product than did wild-type mRNA. In other experiments, this difference was less apparent. No significant difference was observed in the products obtained from the rabbit reticulocyte system (2; unpublished data) and the wheat germ system. In contrast, an *E. coli* cell-free protein system programmed with mRNA from CR17 yielded about 10-fold less gene 0.3 protein than when programmed with wild-type mRNA (6).

To determine the site of initiation for gene 0.3 protein synthesized in vitro, radioactively labeled gene 0.3 protein was subjected to amino-terminal sequence analysis. Because gene 0.3 protein and globin have very similar apparent sizes, and because reticulocyte lysates have high endogenous levels of many amino acids, the wheat germ system was used for experiments involving protein sequence analysis. To prevent amino-terminal acetylation of gene 0.3 protein by the wheat germ system, reactions (0.1 to 0.5 ml) were preincubated with citrate synthetase and oxaloacetate to remove endogenous acetyl coenzyme A (15). The reaction mixtures were then incubated with CR17 or wild-type mRNA, the desired radiolabeled amino acid, and 19 unlabeled amino acids for 1 h at 30°C. Reaction mixtures from syntheses labeled with [^{35}S]methionine and one [^{3}H]-labeled amino acid were combined and supplemented with a small amount of authentic gene 0.3 protein; the gene 0.3 protein was then purified from this mixture by DE52 chromatography at pH 8.0 by using a gradient of NH_4Cl for elution (13). Radiolabeled and authentic 0.3 protein were found to coelute from DE52 at about 0.6 M NH_4Cl. After trichloroacetic acid precipitation, the purified protein was dissolved in formic acid and applied to a Beckman 890C automated protein sequencer. In vitro-synthesized gene 0.3 protein eluted from the DE52 column was radiochemically pure as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and by amino-terminal protein sequence analysis (see below).

Gene 0.3 protein synthesized in a wheat germ extract programmed with wild-type mRNA (or mRNA from the Shine-Dalgarno mutant CR35 [data not shown]) gave an unambigious unique sequence identical to that found for authentic 0.3 protein (Fig. 3; 7). Methionine was found at residues 2, 5, and 17 (Fig. 3B), and tyrosine was found at residues 7, 15, and 24 (Fig. 3D). Thus, initiation begins at the proximal AUG in gene 0.3 mRNA, but the initiating methionine is removed to yield amino-terminal alanine. No hint of a second sequence was observed when synthesis was programmed with wild-type mRNA, indicating that initiation on this nonhomologous mRNA is precise and that removal of the initiating methionine is complete. Initial yields from the sequencer were about 70% of that theoretically expected; in our experience, this yield is average for unblocked proteins prepared as described above.

When mRNA from mutant CR17 was used to program protein synthesis, substantially different results were obtained. The major yield of methionine occurred at residues 3 and 15, and tyrosine was found at residues 5 and 13 (Fig. 3A

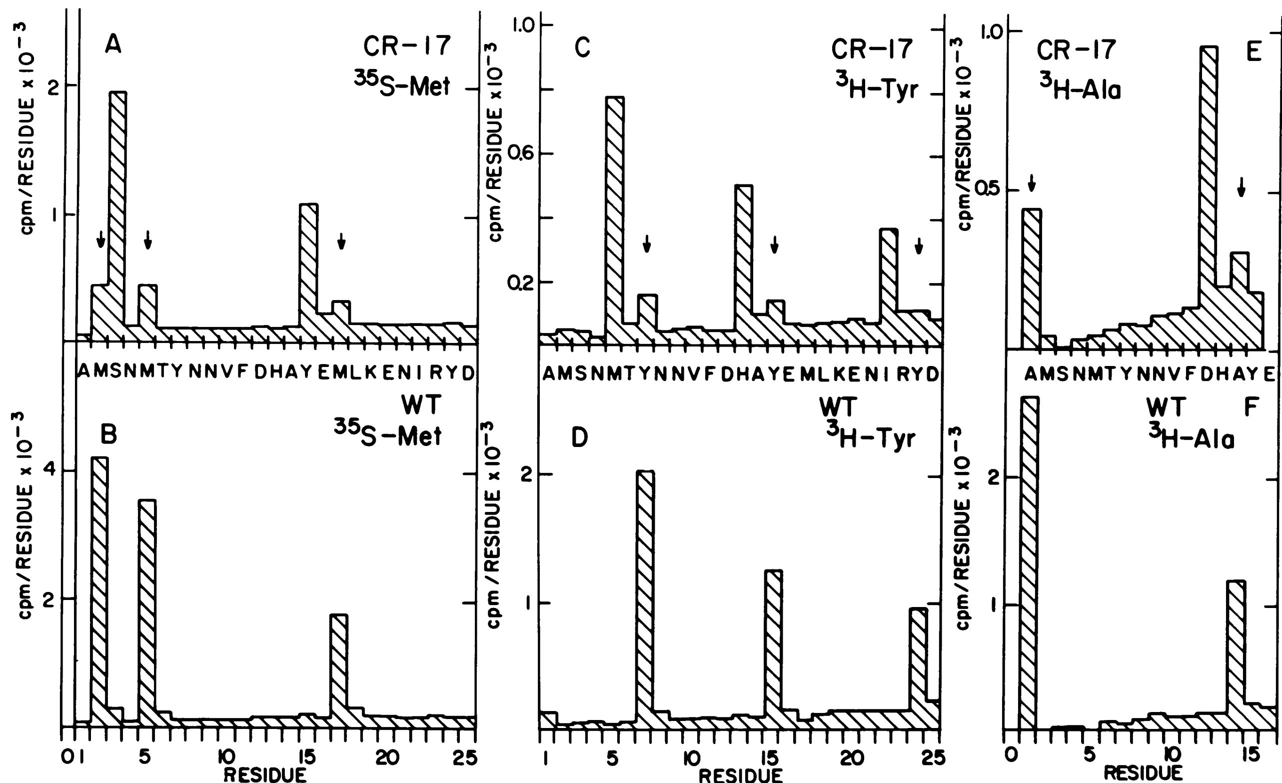

FIG. 3. Amino-terminal sequence analysis of purified T7 gene 0.3 protein synthesized in a wheat germ cell extract programmed with mRNA from wild-type (WT) and mutant CR17 phage. T7 gene 0.3 protein, labeled with [^{35}S] methionine and either [^{3}H]tyrosine or [^{3}H]alanine, was synthesized in wheat germ extracts preincubated with citrate synthetase and oxaloacetate to prevent amino-terminal acetylation (see the text). RNase III-processed early T7 mRNA was produced from WT phage DNA or from the DNA of mutant CR17. T7 0.3 protein was purified from translation extracts by DEAE-cellulose chromatography. Samples were applied to a Beckman 890C protein sequencer, together with apomyoglobin (as carrier protein) and polybrene (to reduce sample washout) as described previously (1). Shown is the amount of radioactivity released after each Edman degradation cycle. For the alanine-labeled samples (panels E and F), the first cycle (cycle 0) was without the cleavage reagent, heptafluorobutyric acid. The sequence of the WT 0.3 protein (7) is given at the top of each lower panel; the position of radioactivity released from initiation at the normal 0.3 protein initiation site (mutated to ACG in CR17) is marked in the upper panel by arrows. The radioactivity applied to the sequencer was as follows. (A) CR17: 27,500 cpm of [^{35}S]methionine. (B) WT: 46,300 cpm of [^{35}S]methionine. (C) CR17: 19,400 cpm of [^{3}H]tyrosine. (D) WT: 28,400 cpm of [^{3}H]tyrosine. (E) CR17: 28,500 cpm of [^{3}H]alanine. (F) WT: 31,600 cpm of [^{3}H]alanine.

and 3C). These results show that when the normal initiation codon was altered to ACG, initiation occurred primarily at the new proximal AUG codon (in this case located two codons farther into the mRNA). Not unexpectedly, the initiating methionine was quantitatively removed from this new amino terminus. It should be noted, however, that with CR17-derived gene 0.3 protein, a small amount of radioactivity was observed at positions expected for wild-type 0.3 protein (Fig. 3A and 3C [arrows]). The yield of radioactivity at wild-type positions represents about 15% of the observed 0.3 protein chains initiated from CR17 mRNA and can be explained only by the initiation of protein synthesis at the position of the ACG (threonine) codon. This level of initiation at the ACG codon (i.e., 12 to 18%) was observed in a number of independent translations and with several different mRNA preparations. Genetic analysis of the phage stock used to make the CR17 mRNA (data not shown) showed that the fraction of wild-type revertants was well less than 0.001%. Thus, we concluded that initiation must have occurred at the ACG codon.

To confirm this result, we prepared gene 0.3 protein labeled with [^{3}H]alanine. As can be seen from Fig. 3E, the small fraction of apparently wild-type 0.3 protein produced with CR17 mRNA had alanine as its first residue, as predicted from the sequence of wild-type gene 0.3 protein. Thus, it seems probable that, in the context of the 0.3 gene mRNA initiation site, an ACG codon can be recognized, albeit at low efficiency, by the initiation species of wheat germ methionine-accepting tRNA. Because of the low yield of the CR17 wild-type gene 0.3 product, we have thus far been unable to prove that $tRNA_i^{Met}$, and not another tRNA, is the species responsible for initiation at the ACG codon.

We cannot be certain that the observed initiation at ACG is not an artifact of in vitro protein synthesis. Furthermore, we used an uncapped heterologous mRNA for these studies. Nevertheless, analysis of the pattern of proteins synthesized in response to early T7 mRNAs and sequence analysis of gene 0.3 protein produced with wild-type mRNA provides strong evidence for a high degree of fidelity in this translation system. With wild-type mRNA, only authentic gene 0.3 protein was synthesized. We estimate that if either of the two AUG codons located immediately distal to the normal 0.3 gene initiation codon (Fig. 1) had served as initiation sites, aberrant initiation at 5% the normal rate would have been detected. Likewise, we would have expected to observe the product if initiation occurred upstream, at the inframe ACG codon at nucleotides 20 to 22. Thus, not every ACG codon serves as an initiation codon, even at the 5% level. If initiation at ACG sites obeys the same efficiency rules as those observed for AUG initiation sites (11), only

those ACG codons that occur in the most favored contexts may function at an efficiency detectable by protein sequence analysis.

What biological role might initiation at ACG serve? Roles similar to the one proposed for regulation of the relative rates of AAV-B and AAV-C protein synthesis (4; Becerra et al., in press) seem likely. Why AAV capsid assembly requires two closely related proteins produced in specific relative amounts is not obvious; however, several instances of proteins closely related in sequence performing different functions have recently been observed (e.g., the adenovirus early region 1A proteins [14]). The studies reported here with in vitro protein synthesis support the hypothesis that ACG functions as an initiation codon at which relatively inefficient rates of initiation are desirable. The significant implication of this finding is that an open reading frame lacking an AUG initiation codon cannot be dismissed as a noncoding region without further analysis.

We are greatly indebted to our colleagues John J. Dunn and F. William Studier for advice and for providing bacteriophage T7 and T7 mutants; we are also indebted to the laboratory of Marshall Elzinga for assistance with the Beckman 890C protein sequencer. We thank Sydell Lamb for providing technical assistance.

This research was supported by the U.S. Department of Energy.

LITERATURE CITED

1. **Anderson, C. W.** 1982. Partial sequence determination of metabolically labeled radioactive proteins and peptides, p. 147–167. *In* J. K. Setlow and A. Hollaender (ed.), Genetic engineering: principles and methods, vol. 4. Plenum Publishing Corp., New York.
2. **Anderson, C. W., J. F. Atkins, and J. J. Dunn.** 1976. Bacteriophage T3 and T7 early RNAs are translated by eukaryotic 80S ribosomes: active phage T3 coded *S*-adenosylmethionine cleaving enzyme is synthesized. Proc. Natl. Acad. Sci. USA **73:**2752–2756.
3. **Anderson, C. W., J. W. Straus, and B. S. Dudock.** 1983. Preparation of cell-free protein synthesizing system from wheat germ. Methods Enzymol. **101:**635–644.
4. **Becerra, S. P., J. A. Rose, and C. W. Anderson.** 1985. Determination of the coding regions for adeno-associated virus structural proteins B and C: evidence that protein B is initiated by an ACG codon. UCLA Symp. Mol. Cell. Biol. **30:**383–396.
5. **Buller, R. M. L., and J. A. Rose.** 1978. Characterization of adenovirus-associated virus-induced polypeptides in KB cells. J. Virol. **25:**331–338.
6. **Dunn, J. J., E. Buzash-Pollert, and F. W. Studier.** 1978. Mutations of bacteriophage T7 that affect initiation of synthesis of the gene 0.3 protein. Proc. Natl. Acad. Sci. USA **75:**2741–2745.
7. **Dunn, J. J., M. Elzinga, K.-K. Mark, and F. W. Studier.** 1981. Amino acid sequence of the gene 0.3 protein of bacteriophage T7 and nucleotide sequence of its mRNA. J. Biol. Chem. **256:**2579–2585.
8. **Dunn, J. J., and F. W. Studier.** 1973. Early T7 RNAs are generated by site-specific cleavages. Proc. Natl. Acad. Sci. USA **70:**1559–1563.
9. **Dunn, J. J., and F. W. Studier.** 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. **166:**477–535.
10. **Green, M. R., and R. G. Roeder.** 1980. Transcripts of the adeno-associated virus genome: mapping of the major RNAs. J. Virol. **36:**79–92.
11. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47:**1–45.
12. **Laughlin, C. A., H. Westphal, and B. J. Carter.** 1979. Spliced adeno-associated virus RNA. Proc. Natl. Acad. Sci. USA **76:**5567–5571.
13. **Mark, K.-K., and F. W. Studier.** 1981. Purification of the gene 0.3 protein of bacteriophage T7, an inhibitor of the DNA restriction system of *Escherichia coli*. J. Biol. Chem. **256:**2573–2578.
14. **Montell, C., G. Courtois, C. Eng, and A. Berk.** 1984. Complete transformation by adenovirus 2 requires both E1a proteins. Cell **36:**951–961.
15. **Palmiter, R. D.** 1977. Prevention of NH_2-terminal acetylation of proteins synthesized in cell-free systems. J. Biol. Chem. **252:**8781–8783.
16. **Srisvastava, A., E. W. Lusby, and K. I. Berns.** 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J. Virol. **45:**555–564.
17. **Studier, F. W.** 1975. Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host. J. Mol. Biol. **94:**283–295.
18. **Zitomer, R. S., D. A. Walthall, B. C. Raymond, and C. P. Hollenberg.** 1984. *Saccharomyces cerevisiae* ribosomes recognize non-AUG initiation codons. Mol. Cell. Biol. **4:**1191–1197.