Formylmethionine Codon AUG as an Initiator of Polypeptide Synthesis

Abstract. *Oligonucleotide messengers containing the sequence AUG at or near the 5′ end of the chain stimulate the incorporation of* N-*formylmethionine, but not of free (unformylated) methionine. AUG is considerably more active in this regard than UUG, the other* N-*formylmethionine codon. The coding activity of subsequent in-phase triplets in the 3′ direction from AUG is markedly stimulated by this codon.*

That *N*-formylmethionine may be involved in the initiation of protein synthesis in *Escherichia coli* has been reported by several groups (*1, 2*). In a comparative study of the two known formylmethionine coding triplets, AUG and UUG (*3*), we have found that AUG is the more active of the two (*4*) as judged by the binding assay of Leder and Nirenberg (*5*), particularly at low Mg^{++} concentrations. Moreover, under these conditions AUG functions as a "phase selector," in that it can automatically fix the reading frame of the polynucleotide message (*4*). This conclusion is based on the observation that, when the AUG sequence is incorporated in a longer oligonucleotide chain, it suppresses the binding of all aminoacyl-sRNA's encoded by out-of-phase sequences, but strongly promotes the binding specified by the adjacent 3′-codon. These results suggest that AUG constitutes at least part of the initiation signal in *Escherichia coli* messenger RNA. Evidence supporting this conclusion is provided by the analysis of polypeptides synthesized under the direction of oligonucleotides containing AUG.

The purpose of these experiments was to determine whether the AUG phase selector would also act as an initiation point in peptide synthesis. On the basis of the data on binding (*4*) we expected to find (i) that the translation of a message containing AUG would always begin at that codon, with the incorporation of formylmethionine at the NH_2-terminus of the peptide chain synthesized, (ii) that AUG would function as an initiator regardless of whether it was located at the 5′-end of the chain or internally, and (iii) that the reading of all subsequent in-phase codons in the 3′ direction from AUG would be markedly enhanced by the presence of this codon.

Early attempts to obtain formylmethionine incorporation with messengers containing AUG or UUG at or near the 5′-end of an oligoU chain [such as $ApUpG(pU)_{\overrightarrow{25}}$ $A(pA)_3pUpG(pU)_{\overrightarrow{25}}$, and so forth] were conducted with a partially purified cell-free system from *E. coli* (*6*). These experiments failed to show significant, reproducible incorporation of formylmethionine, the primary product being pure oligophenylalanine. While the reason for this failure is not known, it seemed possible that the purified system lacked a component, or components, required for correct initiation, since RNA from the virus R17 was also completely inactive as a messenger for formylmethionine incorporation. To test this possibility we have developed a crude cell-free incorporation system (S-30) from *E. coli* similar to that described by Capecchi and Gussin (*7*). That this system is capable of correct initiation is indicated by the fact that R17-RNA stimulates synthesis of large amounts of coat protein containing formylmethionine, as measured in the sucrose-gradient sedimentation assay (*2, 7*).

The oligonucleotide messengers and charged sRNA were prepared as described (*4*). Roughly 65 percent of the charged methionine was formylated, with either tritium-labeled or cold formate. The peptide synthesis reaction contained in a final volume of 50 μl: 20 to 40 μg of oligonucleotide messenger (a saturating amount); 30 to 170 μg of charged sRNA [containing either C^{14}-methionine (200 μc/μmole) and H^3-formate (2000 μc/μmole) or C^{14}-methionine, H^3-phenylalanine (1500 μc/μmole), and cold formate; and all other cold amino acids]; 25 μl of S-30 prepared according to Capecchi and Gussin (*7*), from *E. coli* strain A-19 [lacking ribonuclease I (*8*)], and 15 μl of a stock solution containing salts and an energy source. The final concentrations of these components in the reaction mixture were 75 m*M* ammonium acetate, 3 m*M* adenosine triphosphate, 0.2 m*M* guanosine triphosphate, 10 to 14 m*M* magnesium acetate, 8 m*M* creatine phosphate, 75 μg creatine kinase per milliliter and 50 m*M* tris acetate. The *p*H of this stock solution was adjusted to the desired 6.5 or 7.5 by addition of tris base or acetic acid; the *p*H of the reaction mixture was routinely checked, and it did not change during the incubation. These mixtures were incubated at 37°C for 1 minute, a time usually sufficient for complete reaction with the very short messengers employed.

Two procedures were used for the assay of polypeptide products. Assay A: This procedure was used primarily for the simultaneous measurement of H^3-formate and C^{14}-methionine incorporation. The reaction mixtures were immediately diluted with 1 ml of cold buffer solution (containing 100 m*M* ammonium acetate, 20 m*M* magnesium acetate, and 50 m*M* tris, *p*H 7.2) and layered on 1 ml of a 20-percent sucrose solution, containing the same buffered salts, in a 2-ml centrifuge tube. After centrifugation for 3 hours at 100,000*g*, the ribosomal pellets were resuspended in 50 μl of 0.1*M* KOH and incubated at 37°C for 30 minutes. The solutions were then neutralized with acetic acid and applied to Whatman 3 MM chromatography paper, which was developed in a solvent composed of butanol, glacial acetic acid, and water (45 : 5 : 12.5). Radioactive components were detected in a Vanguard strip scanner, and quantitated by counting paper squares cut from appropriate regions of the chromatogram in a Packard liquid-scintillation system, with a toluene-PPO-POPOP phosphor solution (*4*). In assay B, 2 ml of 5 percent trichloroacetic acid were added to the reaction mixtures, and the resulting suspensions were heated at 100°C for 15 minutes, then cooled, and filtered through Millipore HAWP (cellulose nitrate) filters. The dried filters were then counted in the liquid-scintillation system described above.

The radioactivity (counts/minute) was converted into micromicromoles of amino acid by using predetermined factors. On Millipore filters treated with 5 percent TCA counting efficiencies were 26 percent for C^{14} and 4 percent for tritium, with 40 percent of the C^{14} counts contaminating the tritium counts. With Whatman 3 MM chromatography paper, developed in the above-mentioned solvent, the efficiencies were 25 percent for C^{14} and about 3 percent for tritium, with 50 percent contamination of C^{14} counts in the tritium channel. Although the counting efficiency for tritium on paper varied from one experiment to the next, it could be determined with high precision in experiments in which doubly labeled

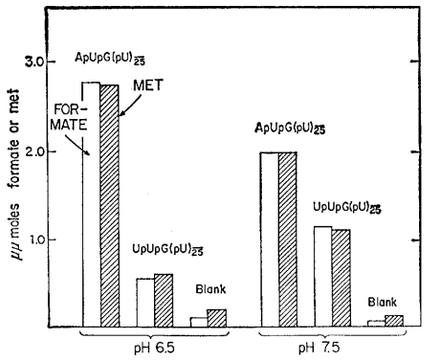

Fig. 1. Incorporation of formyl(H^3)met-(C^{14}) into polypeptide stimulated by ApUpG(pU)$_{25}$ and UpUpG(pU)$_{25}$. The Mg^{++} concentration was 14 mM, and 170 μg of labeled formyl(H^3)met(C^{14})-sRNA was used. The products were analyzed by assay A.

formyl(H^3)methionine(C^{14})-sRNA was employed. This was accomplished by measuring the ratio of the C^{14} to tritium (count/min) in the region of each chromatogram corresponding to unreacted formylmethionine (*9*). This ratio was then used as an internal standard against which the ratio of C^{14} to tritium in the polypeptide regions of the chromatogram could be compared. These ratios of counts could then be converted into micromicromoles without loss of precision.

To determine whether the AUG and UUG codons would direct the incorporation of formylmethionine into polypeptide, the oligonucleotides ApUpG(pU)$_{25}$ and UpUpG(pU)$_{25}$ were used as messengers in the reaction mixture described above, which contained formyl(H^3)methionine(C^{14})-sRNA. Experiments were conducted at *p*H 6.5 and 7.5. Both messengers direct the incorporation of formylmethionine into polypeptide (Fig. 1). Since the ratio of formate to methionine is very close to unity, it may be concluded that virtually no free methionine is incorporated and hence that AUG can be translated at least in this case only as formylmethionine and not as methionine.

Of the two codons AUG is the more active, especially at the lower *p*H (Fig. 1). Thus the decrease in *p*H from 7.5 to 6.5 strongly suppresses the activity of UUG, while that of AUG is slightly increased. This effect of *p*H on the relative coding activities of the two triplets supports our conclusion (*4*) that AUG is a true formylmethionine codon, since Grunberg-Manago and Dondon (*10*) have shown that artifactual miscoding is strongly inhibited at low *p*H.

Having shown that an AUG triplet located at the 5′-end of the chain could be translated unambiguously as formylmethionine, it was now of particular interest to determine whether an internally located AUG would also code for formylmethionine, or whether in this case free methionine would be selected. To answer this question, the oligonucleotides ApUpG(pU)$_{25}$, ApApUpG(pU)$_{25}$ and ApApApUpG(pU)$_{25}$ were used as messengers in the incorporation system (Fig. 2). Both polymers containing an internal AUG sequence code for formylmethionine and not free methionine. This result shows that the basic mechanism of initiation is similar for both internal and 5′-terminal AUG triplets. Moreover, it is entirely consistent with results obtained from binding experiments, which showed that an internal AUG could suppress the reading of out-of phase triplets in the 5′ direction, even if they were located at the 5′-terminal position (*4*). The fact that ApUpG(pU)$_{25}$ is about twice as active as ApApUpG(pU)$_{25}$ or ApApApUpG(pU)$_{25}$ (Fig. 2) was originally interpreted as indicating a preference for initiation at the 5′ terminal triplet. However, a more careful kinetic analysis of the peptide synthesis stimulated by these polymers has recently shown no detectable preference for initiation at the 5′ end of the chain (*11*). It must, therefore, be concluded that if there is such a preference, it is too weak to be detected by experiments of this type.

Further evidence for the initiation function of the formylmethionine codon has been obtained from a study of the influence of AUG on the reading of subsequent triplets in the messenger chain. As pointed out, results from binding experiments (*4*) indicate that the AUG codon should stimulate the translation of all subsequent in-phase codons in the 3′-terminal direction. That this is indeed the case is shown in Fig. 3, where the phenylalanine incorporation directed by ApUpG(pU)$_{25}$ and UpUpGpU$_{25}$ is compared. It is clear that AUG is from 2.5 to 5 times as active as UUG in promoting the reading of subsequent UUU triplets. Moreover, the difference is greatest at the lower Mg++ concentration, where in vitro artifacts are fewest.

When the ratio of phenylalanine to formylmethionine incorporated into

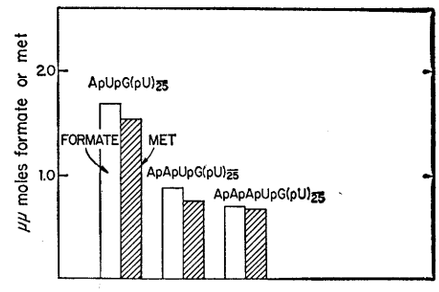

Fig. 2. Incorporation of formyl(H^3)met-(C^{14}) into polypeptide stimulated by ApUpG(pU)$_{25}$, ApApUpG(pU)$_{25}$ and ApApApUpG(pU)$_{25}$. The *p*H of the reaction mixture was 6.5. Products were analyzed by assay A. Blank values of 0.13 $\mu\mu$mole of formate (H^3) and 0.47 $\mu\mu$mole of methionine (C^{14}) have been subtracted from values shown.

polypeptide was determined with the messenger ApUpG(pU)$_{25}$, values between 7 and 14 were obtained, depending upon the charged sRNA preparation used. This result indicates that the AUG codon is translated with high efficiency relative to subsequent triplets in the message.

The foregoing data represent the incorporation achieved in a 1-minute reaction, at the time that peptide synthesis is virtually completed with short messengers. In Fig. 4 is shown the kinetics of phenylalanine and methionine incorporation (the latter being formylated with cold formate) under the direction of the messengers ApUpG(pU)$_{25}$ and UpUpG(pU)$_{25}$. With the polymer containing AUG there is a very rapid burst of synthesis after an initial 5-second lag, which begins to level off after 25 seconds. In contrast, synthesis with UpUpG(pU)$_{25}$ proceeds very slowly, with a gradual increase in rate. Thus while the final amounts of incorporation would appear to indicate that ApUpG(pU)$_{25}$ is about 3.5 times as active as UpUpG(pU)$_{25}$, the maxi-

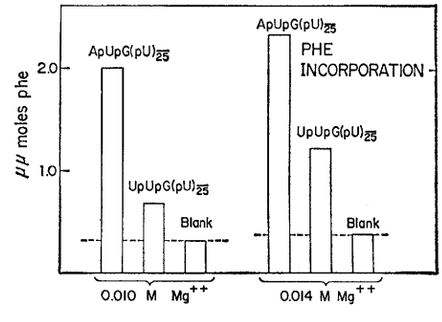

Fig. 3. Incorporation of phenylalanine(H^3) into polypeptide stimulated by ApUpG(pU)$_{25}$ and UpUpG(pU)$_{25}$. The *p*H of the reaction mixture was 6.5, and 30 μg of phe-(H^3)-sRNA were used. Products were analyzed by assay B.

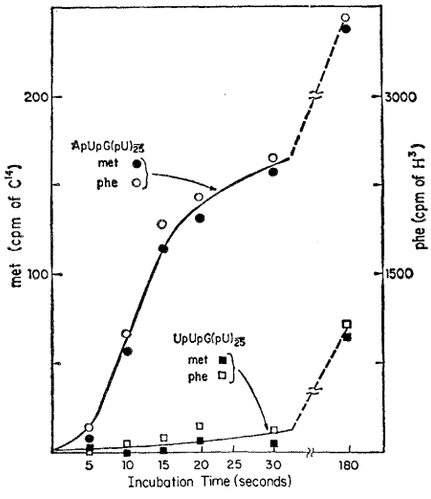

Fig. 4. Kinetics of incorporation of phe(H^3) and met(C^{14}) into polypeptide stimulated by ApUpG(pU)$_{\overline{25}}$ and UpUpG(pU)$_{\overline{25}}$. Each set of points represents a reaction mixture as described, conducted at *p*H 7.0, with 12 m*M* of Mg^{++} and 48 μg of formyl(C^{12})met(C^{14})phe(H^3)-sRNA. Products were analyzed by assay B. Backgrounds of 24 count/min for C^{14}(met) and 135 count/min H^3(phe) have been subtracted from all values shown.

mum rates of synthesis attained show a 20-fold difference between the two messengers. Recent studies with oligonucleotides containing a variety of different codons at the 5′-end of the chain, such as ApCpG(pU)$_{\overline{25}}$, ApApA(pU)$_{\overline{25}}$, and U(pU)$_{\overline{27}}$, suggest that the kinetic pattern observed with UpUpG(pU)$_{\overline{25}}$ is typical of most messengers lacking the AUG sequence (*11*). These results suggest that the activity of the AUG codon as an initiator may be sufficiently strong to account for a large measure of the specificity required of an initiation signal in natural messenger RNA.

The results of our experiments show that AUG can be read unambiguously as formylmethionine; moreover, it is considerably more active in this regard than UUG. Thus we conclude that AUG is a bona fide formylmethionine codon. However, AUG can also code for free methionine, since the repeating triplet polynucleotide $(pApUpG)_n$ stimulates the synthesis of polymethionine (*12*). In order to reconcile these two apparently contradictory results, we propose that the AUG codon can be read in two different ways. (i) It may be read as the initial triplet in the message, in which case formylmethionine will be incorporated; or (ii) if encountered during the normal process of messenger readout, it may be read as an internal codon so that free methionine will be incorporated. The physical basis for the distinction between these two different modes of reading is not yet known. However, as suggested (*4, 13*), the specificity of the two sRNA binding sites on the ribosome (*17, 18*) may be involved in determining which of the two aminoacyl-sRNA's (formylmethionine- or methionine-sRNA) will be bound. Thus in the process of initiation, if the AUG sequence first appeared at the peptidyl-sRNA site, formylmethionine-sRNA (which resembles peptidyl-sRNA in that the α-amino group is involved in an amide bond) would be specifically bound. Alternatively, in the process of messenger readout the peptidyl-sRNA site would be permanently occupied, and hence an AUG triplet would automatically appear first at the aminoacyl-sRNA site; in this case methionine-sRNA would be selected.

The question of whether the natural initiation site on the messenger RNA consists simply of an AUG codon, or whether other specific factors are also involved, remains to be explored. However, it does seem likely that additional control elements of some kind must be present in order to prevent mistaken initiation at internal AUG's which were originally intended to be read as free methionine codons. The nature of these additional control elements is unknown (*4*).

Regardless of whether AUG is sufficient to specify an initiator site in vivo, the mechanism by which it stimulates initiation with synthetic messengers in vitro is of interest. The AUG sequence constitutes a preferential binding site in the absence of protein synthesis; for this reason it has been termed a phase-selector codon (*4*). This result suggests that the high rate of initiation at AUG may be due simply to its ability to stimulate the rapid formation of a strong active complex with the ribosome (consisting of messenger, formylmethionine-sRNA, aminoacyl-sRNA corresponding to the codon adjacent to AUG and 70*S* ribosome). However, the results of the kinetic analysis shown in Fig. 4 indicate that the initiator capacity of AUG may be too high to be accounted for by binding ability alone. Thus when compared in binding experiments AUG was at best only eight times more active than UUG (*4*), whereas it now seems that there may be as much as a 20-fold difference on the basis of rates of peptide synthesis.

In contrast to the satisfactory results obtained with the S-30 incorporation system is the failure of a more purified system (*6*) to initiate properly. It seems likely that this failure may be due to the absence of a component which is required for specific initiation. What the chemical nature or mode of action of this component might be is unknown. However, it would appear to be involved at some stage subsequent to the formation of the tetramolecular complex of messenger, ribosome, formylmethionine-sRNA and aminoacyl-sRNA, since binding experiments have shown that well-washed ribosomes are quite active in this regard (*4*).

R. E. Thach, K. F. Dewey
J. C. Brown, Paul Doty
Department of Chemistry, Harvard University, Cambridge, Massachusetts

References and Notes

1. J. P. Waller, *J. Mol. Biol.* **7**, 483 (1963); B. F. C. Clark and K. A. Marcker, *Nature* **207**, 1038 (1965); R. E. Webster, D. L. Engelhardt, N. Zinder, *Proc. Nat. Acad. Sci. U.S.* **55**, 155 (1966); T. Nakamoto and D. Kolakofsky, *ibid.*, p. 606.
2. J. M. Adams and M. R. Capecchi, *Proc. Nat. Acad. Sci. U.S.* **55**, 147 (1966).
3. Conventions and abbreviations used: pA, pU, pC, and pG represent adenylic, uridylic, cytidylic, and guanylic residues, respectively; the letters A, U, C, and G are used to designate the same nucleotides when referring to a general coding sequence, as opposed to a specific chemical compound (for example, AUG represents a coding sequence, while ApUpG represents a trinucleoside diphosphate); oligonucleotide sequences are written with the 5′-end on the left, the 3′-end on the right; numerical subscripts designate the number of a given type of residue in sequence, a bar over the subscript signifying a number average; formylmet, and phe stand for *N*-formylmethionine and phenylalanine, respectively; cpm stands for counts per minute; S-30 represents a supernatant fraction, prepared from a crude cell-free extract of *E. coli*, strain A-19, obtained after centrifugation at 30,000*g*; sRNA, soluble RNA.
4. T. A. Sundararajan and R. E. Thach, *J. Mol. Biol.*, in press.
5. M. W. Nirenberg and P. Leder, *Science* **145**, 1399 (1964).
6. R. E. Thach, *et al.*, *Proc. Nat. Acad. Sci. U.S.* **54**, 1167 (1965).
7. M. R. Capecchi and G. Gussin, *Science* **149**, 417 (1965).
8. R. F. Gesteland, *J. Mol. Biol.* **16**, 67 (1966).
9. A small amount of unreacted formylmethionine was always found in chromatograms of reactions containing ApUpG(pU)$_{\overline{25}}$. This presumably represents formylmethionine-sRNA which is bound to ribosomes but not incorporated into polypeptide. The ratio of C^{14} to tritium in this formylmethionine component was identical to that of authentic formylmethionine obtained from a chromatogram of an alkaline hydrolyzate of formyl-(H^3)methionine(C^{14})-sRNA.
10. M. Grunberg-Manago and J. Dondon, *Biochem. Biophys. Res. Commun.* **18**, 517 (1965).
11. J. C. Brown and R. E. Thach, manuscript in preparation.
12. R. D. Wells, T. M. Jacob, H. R. Kossel, A. R. Morgan, S. A. Narang, E. Ohtsuka, H. G. Khorana, 50th Ann. Meeting Fed. Amer. Soc. Exp. Biol., April 1966.
13. H. Nöll, *Science* **151**, 1241 (1966).
14. J. R. Warner and A. Rich, *Proc. Nat. Acad. Sci. U.S.* **51**, 1134 (1964).
15. H. Kaji and A. Kaji, *ibid.* **52**, 1541 (1964).
16. Supported by NIH grant HD-01229 and NSF grant GB-4563.

10 May 1966