NUCLEOTIDE SEQUENCE OF A RIBOSOME ATTACHMENT SITE OF BACTERIOPHAGE f2 RNA

S. L. Gupta, J. Chen, L. Schaefer, P. Lengyel and S. M. Weissman
Department of Molecular Biophysics and Biochemistry and the
Department of Medicine, Yale University, New Haven, Connecticut 06520

Received April 21, 1970

Summary. The incubation of f2 bacteriophage RNA with E. coli ribosomes, initiation factors, GTP, and formylmethionyl-$tRNA_F$ results in the attachment of a ribosome to the coat protein initiation site of the f2 RNA. Added ribonuclease T_1 degrades all parts of the f2 RNA except that protected by the attached ribosome. The sequence of the protected fragment was established: it specifies the first 7 amino acids of the coat protein and extends 41 nucleotides towards the 5' end from the AUG initiator codon. There is no terminator codon directly adjacent to the 5' side of the AUG codon. The f2 coat initiation site as well as three of the four other initiation sites whose sequence is known include a UUUGA sequence which is separated from the AUG codon by 2 or 3 nucleotides. It is a curious fact that the sequence of the TψC containing loop of the initiator $tRNA_F^{Met}$ is complementary to this common sequence.

The initiator of most (if not all) of the proteins in E. coli is formylmethionyl-$tRNA_F$ (fMet-$tRNA_F$) (for a review see ref. 1.). The codons specifying fMet-$tRNA_F$ in the in vitro system are AUG and GUG (initiator codons). The following facts indicate that an initiator codon is not the entire initiation signal: in addition to specifying fMet-$tRNA_F$, AUG and GUG also specify internal aminoacyl residues of the peptide chain (AUG may stand for Met-$tRNA_M$ and GUG for Val-tRNA) and AUG and GUG sequences also occur in mRNA as overlaps of adjacent codons. Initiation signals are not restricted to the 5' end of mRNA (1). Not every initiation signal is recognized by every ribosome-initiation factor complex (2) and the frequency of initiation at different initiation sites can be vastly different (3).

The features which could in principle determine if an AUG or GUG codon is recognized as an initiator codon and if so determine its efficiency include the sequence of nucleotides at the 5' side of the codon and the secondary structure of a messenger. To contribute to the solution of this problem several laboratories have examined the nucleotide sequence of initiation sites in the RNA of RNA bacteriophages (eg. R17, Q_β , f2) (4-7). This RNA serves as a messenge

which directs the synthesis in infected E. coli or in a cell-free system of at least three proteins: phage coat protein, replicase enzyme and maturation protein (3 et op. cit.).

To be able to sequence the initiation site without having to sequence the entire f2 RNA (approx. 3300 nucleotides) we had developed a procedure for isolating the initiation site containing segment of the f2 RNA. We established that in the presence of fMet-$tRNA_F$ (as the only aminoacyl-tRNA), GTP, and initiation factors, a ribosome binds to f2 RNA forming a 70S initiation complex (8). Ribosomes were known to protect the mRNA region to which they are bound against cleavage by nucleases (9). Consequently, we formed a 70S initiation complex, treated it with ribonuclease T_1, and isolated the f2 RNA fragment which was protected against cleavage. In a previous communication we described that this fragment of f2 RNA is a unique segment at the center of which is the AUG codon specifying the initiation of the synthesis of the phage coat protein and moreover, that none of the known terminator codons (UAA, UAG, UGA) is adjacent to the 5' side of this initiator codon (5).

In this communication we present the nucleotide sequence of the entire 61 nucleotide long fragment (Figure 1). The sequence and amount of the various oligonucleotides obtained by digesting the fragment with ribonuclease T_1 and pancreatic ribonuclease is shown in Table 1.

The nucleotide sequence proceeding in the 3' direction from the AUG codon specifies the first seven amino acids of the f2 coat protein (10). This proves that the fragment contains the coat initiation site. The lack of a terminator codon directly adjacent to the 5' side of the initiator codon seems to indicate that there are untranslated parts on the f2 RNA. The fragment extends 41 nucleotides in the 5' direction from the AUG codon. The nucleotide sequence of one peptide chain termination signal has recently been established by Nichols (11) as UAAUAG. The lack of this hexanucleotide in the 41 nucleotide long segment preceding the initiator AUG is consistent with the possibility that no part of this sequence is translated. The sequence of the fragment can be

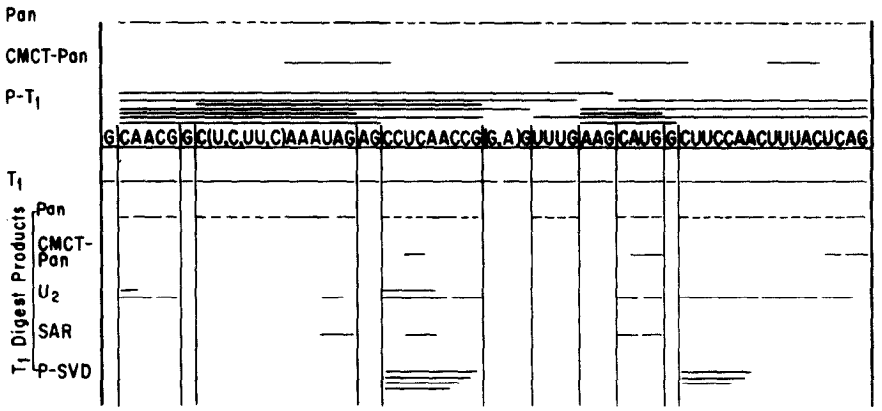

Figure 1. Nucleotide sequence of a ribosome attachment site of bacteriophage f2 RNA. ^{32}P-labelled f2 RNA was incubated in a reaction mixture including ribosomes, initiation factors, GTP, salts, and as the only aminoacyl-tRNA: fMet-tRNA$_F$ (5). During the incubation part of the f2 RNA became bound to ribosomes (8). Subsequently the reaction mixture was incubated with added ribonuclease T_1 in order to digest all regions of the f2 RNA except that protected by the ribosome bound to it. The ribosomes with the protected fragment were separated by centrifugation through a sucrose gradient from the products of digestion of those regions of the f2 RNA which were not bound to ribosomes. The fractions containing the ribosomes with the protected fragment were pooled and the RNA isolated by phenol extraction (5). The fragment was isolated by acrylamide gel electrophoresis (5) and the nucleotide sequence of the fragment was established by enzymatic digestion of the fragment with various nucleases and the characterization of the cleavage products by electrophoresis on DEAE cellulose paper at pH 1.7 and 3.5, base analysis and further enzymatic digestion according to published procedures (16,17,18,19,20). The sequences of the segments enclosed in parentheses are tentative. Bars placed over the sequence delimit segments obtained by: a) digestion of the fragment with pancreatic ribonuclease (Pan), b) partial modification of the fragment with N-cyclohexyl-N'-(β morpholinyl-4-0-methyl)-carbodiimide-methyl-p-toluene sulphonate (CMCT) followed by digestion with pancreatic ribonuclease (CMCT-Pan), c) partial digestion of the fragment with ribonuclease T_1 (P-T_1). Bars placed under the sequence delimit segments obtained by: d) extensive digestion of the fragment with ribonuclease T_1 (T_1). (These segments are also delimited by vertical lines:) e) further digestion of the individual segments produced in the extensive digestion with ribonuclease T_1 (T_1 digest products) with I) pancreatic ribonuclease (Pan), II) pancreatic ribonuclease after extensive modification with CMCT (CMCT-Pan), (21,22), III) U_2 ribonuclease (U_2) (23), IV) spleen acid ribonuclease (SAR), and V) snake venom diesterase (P-SVD). It was generally assumed that: products of T_1 ribonuclease digestion have a 3' terminal Gp; products of pancreatic ribonuclease digestion have a 3' terminal Up or Cp; products of U_2 ribonuclease digestion have a 3' terminal Ap or Gp; products of the digestion of the segment modified by CMCT with pancreatic ribonuclease have a 3' terminal Cp; snake venom diesterase removes mononucleotides from the 3' terminal; and spleen acid ribonuclease has a complex pattern of specificity (18,20).

---

arranged in several extensively hydrogen bonded structures (see also 4,6,7).

The physiological relevance of this is however not known.

Table 1. Amounts of various products in enzymatic digests of a ribosome attachment site of f2 RNA. The amount of UUUG was taken as 1 mole and the amount of the other products was related to that of UUUG. (T_1), ribonuclease T_1; the relative amounts shown are the average of 6 experiments. (Pan), pancreatic ribonuclease; the relative amounts shown are the average of 8 experiments. For details see the legend to Figure 1. Spots 102A and 102B were not well separated on the two dimensional electropherogram on DEAE cellulose paper. The materials in the two spots were separated from each other by homochromatography on DEAE cellulose thin layer plates (24).

Enzyme Used	Spot number on auto-radiogram (5)	Sequence of product	Relative Amount of Product	Integer used in constructing the proposed sequence of the site
	1	G	3.08	3
	3	AG	1.98	2
	8	AAG	0.89	1
	28	CAUG	1.13	1
	40	UUUG	1.00	1
(T_1)	102A	CUUCCAACUUUACUCAG	0.73	1
	102B	C(U,C,UU,C) AAAUAG	0.79	1
	103	CCUCAACCG	0.83	1
	105	CAACG	0.79	1
	1	C	8.50	9
	2	U	10.70	12
	3	AC	1.15	1
	4	AU	1.15	1
	7	AAC	3.00	3
(Pan)	13	GGC	1.43	2
	16	AAAU	1.02	1
	54	G(G,A) GU	0.83	1
	106	AGAGC	0.96	1
	107	GAAGC	0.99	1
	108	AG	0.80	1

Bacteriophage f2 is closely related to bacteriophage R17. The nucleotide sequence of three initiation sites in R17 was recently determined by J. Steitz (4,6). She used techniques similar to those described here except that she employed pancreatic ribonuclease to prepare a 39 nucleotide long fragment containing the coat initiation site. There is an excellent agreement between the sequences proposed for the f2 and the R17 coat initiation sites: (4,6, see also 12). Only two nucleotides are different out of 39.

The conditions of the preparation and nuclease treatment of the f2 RNA and R17 RNA initiation complexes were somewhat different. Steitz obtained

Table 2. Nucleotide sequences of (a) initiation sites in *E. coli* RNA phages and (b) of the TψC containing loop of *E. coli* $tRNA_F^{Met}$. The sequences of the initiation sites are printed in the 5' (left) to 3'(right) direction, the $tRNA_F^{Met}$ sequence is shown in the 3' (left) to 5' (right) direction. The AUG sequence of the initiation sites specifies the fMet-$tRNA_F$ which initiates the translation of the proteins indicated. The sequences which are complementary to the sequence of the TψC containing loop in $tRNA_F^{Met}$ are underlined.

	Bacteriophage	Ribosome attachment site	Sequence	Reference
a)	R17	replicase	GGAUUACCCAUG	(4,6)
	R17	maturation protein	GUUUGACCUAUG	(4,6)
	Q_β	coat protein	AAUUUGAUCAUG	(7)
	R17	coat protein	GUUUGAAGCAUG	(4,6)
	f2	coat protein	GUUUGAAGCAUG	this paper
b)	*E. coli* $tRNA_F^{Met}$: TψC containing loop		UAAACψT..	(14)

binding of ribosomes to all three initiation sites, presumably in part as a consequence of the fragmentation of the R17 RNA occurring prior to ribosome binding. Under our conditions less than 5% of the f2 RNA was fragmented before and during ribosome binding, and we detected only a single initiation site. These results are consistent with the view that the replicase initiation site is unavailable for ribosome attachment, unless the RNA is unfolded by either fragmentation or translation of the coat gene and furthermore, the maturation protein initiation site is relatively inefficient (13).

Four of the five known initiation sites (all in phage RNA) include a UUUGA sequence separated from the initiator codon by 2 or 3 nucleotides (Table 2). It is a curious fact that this common UUUGA sequence is complementary (in the anti-parallel direction) to the sequence of the TψC containing loop of the initiator $tRNA_F^{Met}$ (14). ($tRNA_F^{Met}$ is the only tRNA of *E. coli* which is known to have this sequence in the TψC loop (15).) The significance of this complementarity remains to be seen.

This study has been supported by grants from the National Institute of General Medical Sciences, the National Cancer Institute and the American Cancer Society. We thank Professors F. Egami for the U_2 ribonuclease and G. Bernardi for the spleen acid ribonuclease.

REFERENCES

1. Lengyel, P., and Soll, D., Bact. Rev. 33, 264 (1969).
2. Lodish, H. F., Nature 224, 867 (1969).
3. Lodish, H. F., Nature 220, 345 (1968).
4. Steitz, J. A., Cold Spring Harbor Symp. Quant. Biol. 34, 621 (1970).
5. Gupta, S. L., Chen, L., Schaefer, L., Weissman, S. M., and Lengyel, P., Cold Spring Harbor Symp. Quant. Biol. 34, 630 (1969).
6. Steitz, J. A., Nature 224, 957 (1969).
7. Hindley, J., Staples, D. H., Nature 224, 964 (1969).
8. Kondo, M., Eggertsson, G., Eisenstadt, J., and Lengyel, P., Nature 220, 368 (1968).
9. Takanami, M., Yan, Y., and Jukes, T. H., J. Mol. Biol. 12, 761 (1965).
10. Weber, K., and Konigsberg, W., J. Biol. Chem. 242, 3563 (1967).
11. Nichols, J. L., Nature 225, 147 (1970).
12. Robinson, W. E., Frist, R. H., and Kaesberg, P., Science 166, 1291 (1969).
13. Lodish, H. F., and Robertson, H. E., Cold Spring Harbor Symp. Quant. Biol. 34, 655 (1969).
14. Dube, S. K., and Marcker, K. A., European J. Biochem. 8, 256 (1969).
15. Zachau, H. G., Angew. Chem. Internat. Edit. 8, 711 (1969).
16. Sanger, F., Brownlee, G. G., and Barrell, B. G., J. Mol. Biol. 13, 373 (1965).
17. Brownlee, G. G., and Sanger, F., J. Mol. Biol. 23, 337 (1967).
18. Brownlee, G. G., Sanger, F., and Barrell, B. G., J. Mol. Biol. 34, 379 (1968).
19. Forget, B. G., and Weissman, S. M., J. Biol. Chem. 243, 5709 (1968).
20. Forget, B. G., and Weissman, S. M., J. Biol. Chem. 244, 3148 (1969).
21. Gilham, P. T., J. Am. Chem. Soc. 84, 687 (1962).
22. Adams, J. M., Jeppesen, P.G.N., Sanger, F., and Barrell, B. G., Nature 223, 1009 (1969).
23. Arima, T., Uchida, T., and Egami, F., Biochem. J. 106, 609 (1968).
24. Brownlee, G. G., and Sanger, F., Europ. J. Biochem. 11, 395 (1969).