STUDIES ON THE BACTERIOPHAGE MS2. G–U–G AS THE INITIATION CODON OF THE A-PROTEIN CISTRON*

G. VOLCKAERT and W. FIERS

Laboratory of Molecular Biology and Laboratory of Physiological Chemistry, University of Ghent, Belgium†

Received 18 June 1973

1. Introduction

J. Argetsinger-Steitz [1] was able to bind *Escherichia coli* ribosomes specifically to the initiation sites of the three cistrons present in the bacteriophage R17 genome. As the ribosomes protected these regions over a length of approx. 35 nucleotides against nuclease attack, it was possible to isolate them and to determine their sequence. By comparison with the N-terminal amino acids of the respective proteins she could show that all three cistrons were initiated with an A–U–G codon. In the case of bacteriophage MS2, which is very closely related to R17, we have directly sequenced the regions where the coat protein and the replicase are initiated, and these regions were found to be almost identical to their R17 counterparts [2]. The three cistrons of the bacteriophage Qβ, which is not serologically related, also start with A–U–G signals [3–5] as well as an ΦX-174 gene, which was recently identified [6].

We have previously isolated and sequenced a 125 nucleotide fragment derived from the 5′-end of the MS2 RNA chain [7, 8]. As we found an overlap of 16 nucleotides with the ribosome binding region of the cistron which codes for the A-protein [1], we could conclude that this first gene starts at nucleotide 130 from the 5′-end. We have now been able to prolong this MS2 RNA sequence and found that the initiating codon is G–U–G. This result then shows that protein synthesis can start as well at this G–U–G signal as at an A–U–G codon.

2. Methods

MS2 RNA, labelled with ^{32}P, was partially digested with ribonuclease T_1. The resulting digest was fractionated first by electrophoresis on a polyacrylamide gel slab into a number of bands and each band was then further separated by two-dimensional electrophoresis into pure fragments. These procedures have been described in detail previously [7–10], as well as the nomenclature of the fragments [8]. Digestion with carboxymethyl-ribonuclease A (CM-RNAase) has been described by Contreras and Fiers [11]; the conditions used were 20 min incubation at 0°C with an enzyme to substrate ratio of 1:600. Further analysis of the fragments was mainly according to the methods of Sanger et al. [12–14].

3. Results

The untranslated 5′-terminal leader sequence, which we previously published [7, 8], contains near its right end the heptanucleotide A–G–G–A–G–G–U (position 115–122). This heptanucleotide occurs only once in the complete MS2 RNA chain [15], and hence can be used as an indicator for this region. On this basis we have isolated a series of new fragments, summarized in table 1, which all contain this oligonucleotide.

Maps of the fragment $\delta_6 b_5$ after digestion with ribonuclease T_1 (RNA T_1) or pancreatic ribonuclease

* Publication XXVII of this series.

† Mailing address: Prof. Dr. W. Fiers, Laboratorium voor Moleculaire Biologie, Ledeganckstraat, 35, 9000 Gent, België.

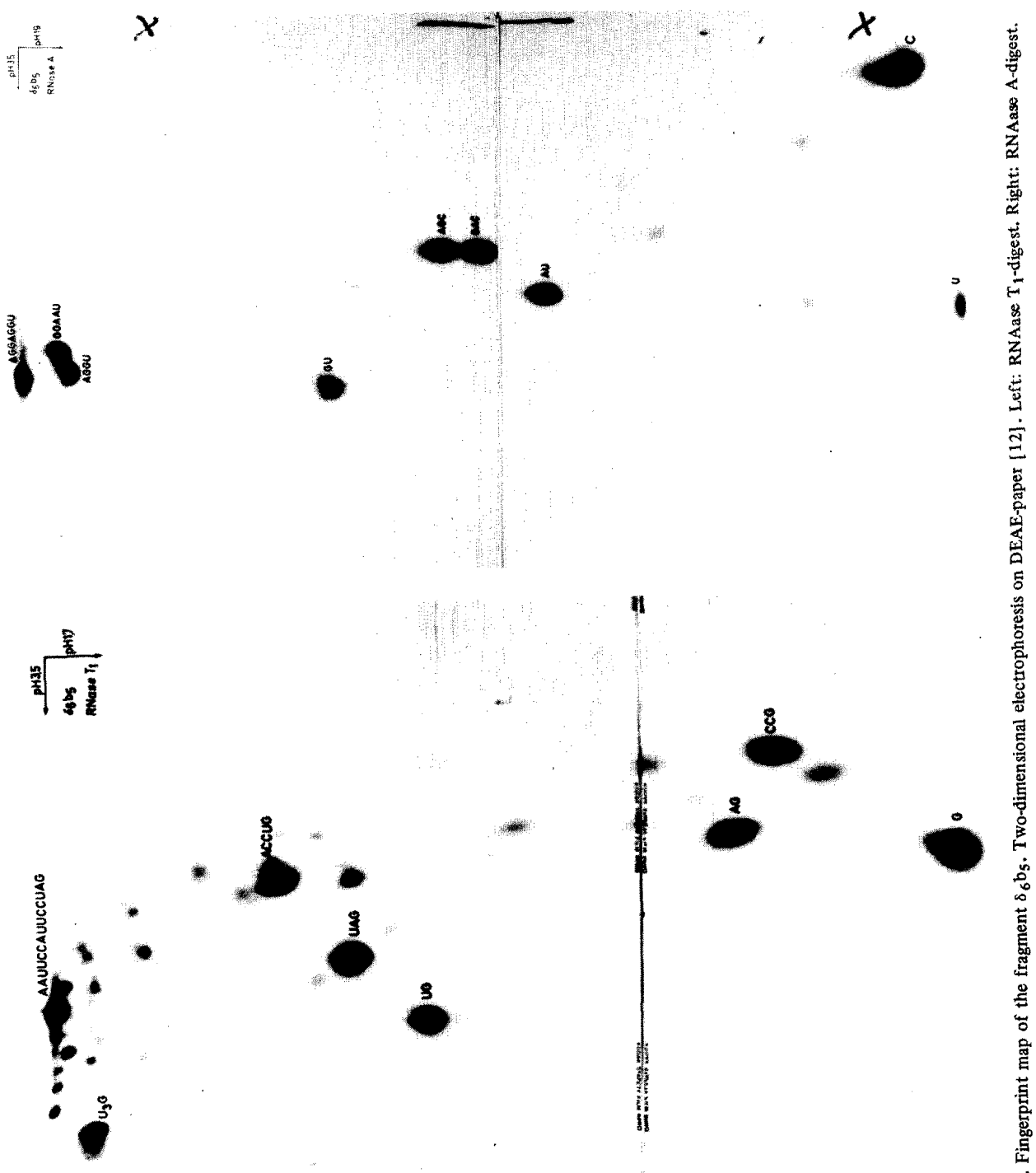

Fig. 1. Fingerprint map of the fragment $\delta_6 b_5$. Two-dimensional electrophoresis on DEAE-paper [12]. Left: RNAase T_1-digest. Right: RNAase A-digest.

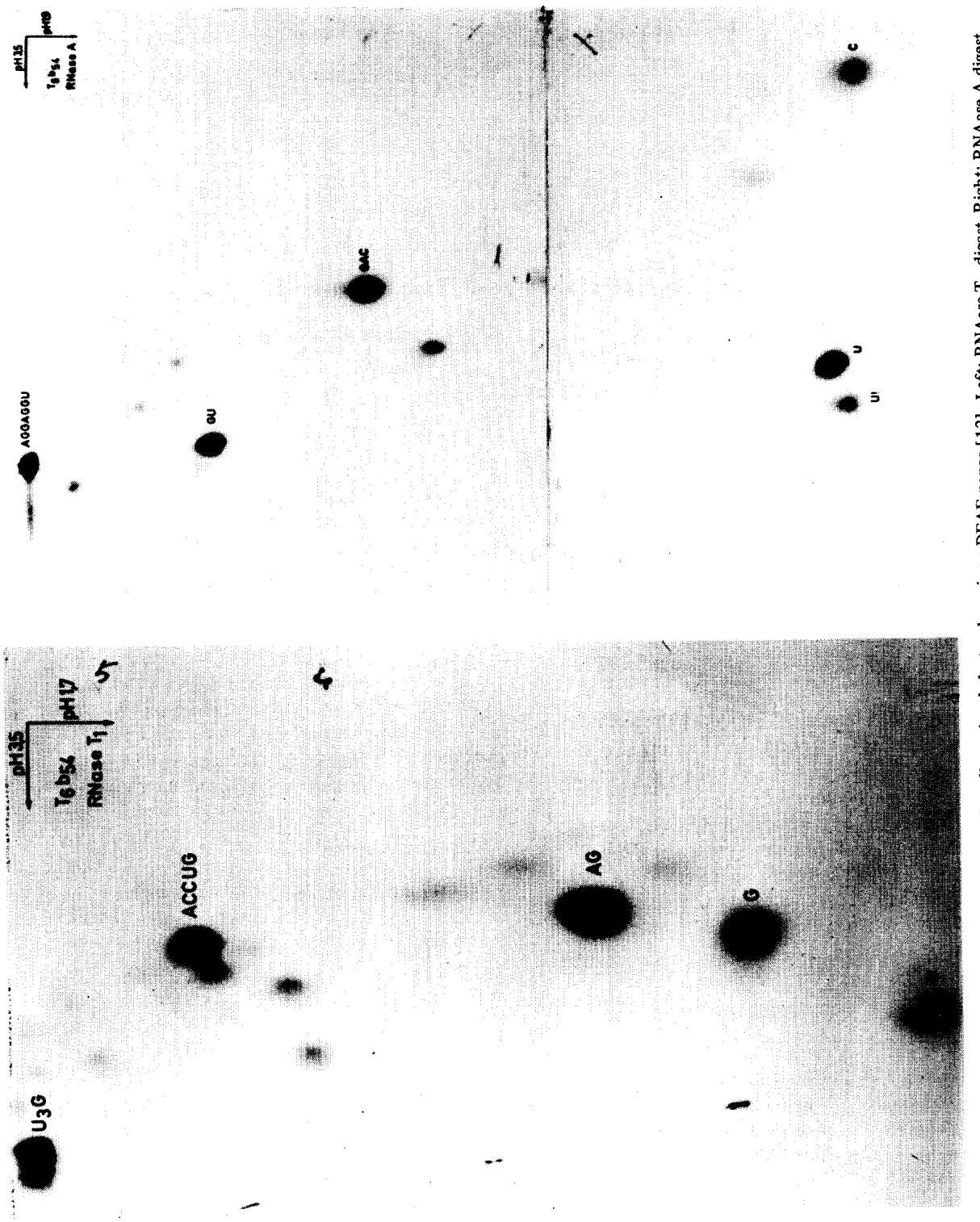

Fig. 2. Fingerprint map of the fragment T_6b_{54}. Two-dimensional electrophoresis on DEAE-paper [12]. Left: RNAase T_1-digest. Right: RNAase A-digest.

Table 1
Fragments derived from the G–U–G initiator region of the A-protein cistron.

Fragment	Position[a]
$\delta_5 b_7$; $\delta_6 b_5$ $\delta_7 b_{10}$; $D_4 b_5$	92–132
$\gamma_4 b_{01}$	25–132
$T_6 b_{54}$[b]	115–131

[a] Number of nucleotides from the 5′-terminus (fig. 3).
[b] Fragment isolated from a CM-RNAase digest of MS2 RNA.

(RNAase A) are shown in fig. 1 and the oligonucleotide composition is reported in table 2. Data for some other fragments were identical (table 1). A comparison of this oligonucleotide composition with the 5′-terminal sequence reported previously [7, 8] reveals the presence of two new oligonucleotides, U–G and A–C–C–U–G in the RNAase T_1 map. The sequence of the former follows from its position and from its composition after alkaline hydrolysis. The sequence of the latter is derived from the following analyses (results expressed as molar yields): Digestion with RNAase A: 1.0 A–C; 1.2 C; 1.4 U; 1.1 G. Digestion with ribonuclease U_2: A and (C_2, U) G. Digestion with RNAase A after blocking with carbodiimide [8]: A–C; C; U–G.

The RNAase A fingerprint contained the additional oligonucleotides G–U and G–A–C (after digestion with RNAase T_1: G and A–C). But these overlaps did not allow an unambiguous ordering.

The fragment $T_6 b_{54}$, however, contains the sequence A–C–C–U–G but not U–G in the RNAase T_1 fingerprint (fig. 2 and table 3); instead the 3′ end is U. The RNAase A overlaps G–A–C and G–U are still present. Therefore the relative order must be as shown in fig. 3.

4. Discussion

The sequence of the first 132 nucleotides of MS2 RNA are shown in fig. 3 in a secondary structure model. The 20 nucleotides preceding the initiator A–U–G codon on the R17 A protein ribosomal binding site [1] are also present in MS2 RNA. In addition, the sequence of R17 RNA up to position 117 [16] was found to be identical to its MS2 RNA counterpart. Furthermore, the nucleotide sequence following the G–U–G is also identical to the one in R17 (our unpublished results) and the first 5 amino acids in the MS2 A-protein are the same as those reported for the

Table 2
Oligonucleotide composition of fragment $\delta_6 b_5$.

RNAase T_1 oligonucleotides[a]		Molar equivalents Found	Theoretical
0	G	5.3	4
001	AG	1.6	1
020	CCG	1.0	1
100	UG	1.2	1
101	UAG	1.8	2
121	ACCUG	1.0	1
300	UUUG	1.0	1
544	AAUUCCAUUCCUAG	0.9	1
RNAase A oligonucleotides			
	U	9.4	7
	C	7.7	6
	G	1.0	1
	AU	1.3	1
	GU	0.9	1
	GAC	1.1	1
	AGC	1.3	1
	AGGU	1.0	1
	GGAAU	1.2	1
	AGGAGGU	0.8	1

[a] A standard numbering of these oligonucleotides was adopted [8].

Table 3
Oligonucleotide composition of fragment $T_6 b_{54}$.

RNAase T_1 oligonucleotides		Molar equivalents Found	Theoretical
0	G	2.8	2
	U	0.8	1
001	AG	2.2	2
121	ACCUG	0.8	1
300	UUUG	1.0	1
RNAase A oligonucleotides			
	U	3.3	3
	C	1.4	1
	GU	1.2	1
	GAC	0.9	1
	AGGAGGU	0.9	1

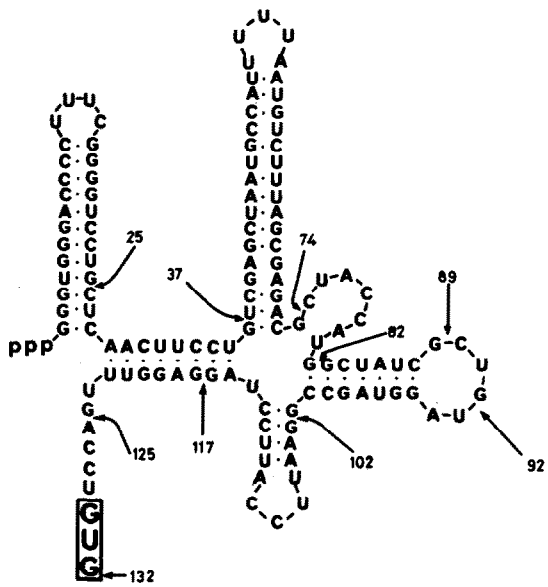

Fig. 3. Secondary structure of the 5'-terminal region of MS2 RNA. This is an extension of the model published previously [8] (slightly improved version). The arrows point to sites easily split by RNAase T_1. The A-protein gene starts at position 130 from the 5'-terminus.

R17 protein (Vandekerckhove and Van Montagu, personal communication; [17]).

This strong similarity to the R17 ribosome binding site allows us to conclude that in MS2 RNA G–U–G is indeed used as an initiator codon for the A-protein gene. It may be noted that Revel et al. [18] isolated directly ribosomal binding sites from MS2 RNA. On their published maps, one can detect both A–C–C–U–A–U–G and A–C–C–U–G; in retrospect it seems likely that their stock of MS2 virus was heterogeneous and contained two forms of viral RNA.

R17 and MS2 are nearly identical phages in all molecular biological aspects, including the level of *in vivo* A-protein synthesis and the recognition of the A-protein ribosomal binding region in the *in vitro Bacillus stearothermophilus* system. Therefore it seems that A–U–G and G–U–G are nearly equally efficient initiator codons. Although due to secondary structure, the A-protein initiator region is normally not available [1, 19], Argetsinger-Steitz [20] has recently shown that this region is intrinsically very efficient in stimulating polypeptide synthesis.

Our results on the occurrence of G–U–G as an initiator codon in a natural genome nicely confirm the original findings of Clark and Marcker [21] and of Ghosh et al. [22] who showed that ribosome-mediated binding of formyl-met-tRNA took place not only with the trinucleotide A–U–G but also with G–U–G. They further demonstrated that G–U–G, present in a polymer, could direct the initiation of a polypeptide starting with formyl-methionine. On the other hand, Stewart et al. [23] presented evidence indicating that in a yeast messenger RNA a G–U–G codon could not function *in vivo* as a signal for initiation of translation.

Acknowledgements

Research was supported by the Belgian Fonds voor Fundamenteel Wetenschappelijk Onderzoek (grant no. 841). G.V. thanks the I.W.O.N.L. for a fellowship. We are grateful to our coworkers for gifts of materials and for information.

References

[1] Argetsinger-Steitz, J. (1969) Nature 224, 957.
[2] Min Jou, W., Haegeman, G., Ysebaert, M. and Fiers, W. (1972) Nature 237, 82.
[3] Staples, D.H., Hindley, J., Billeter, M.A. and Weissmann, C. (1971) Nature New Biol. 234, 202.
[4] Staples, D.H. and Hindley, J. (1971) Nature New Biol. 234, 211.
[5] Argetsinger-Steitz, J. (1972) Nature New Biol. 236, 71.
[6] Robertson, M.D., Barrell, B.G., Weith, H.L. and Donelson, J.E. (1973) Nature New Biol. 241, 38.
[7] De Wachter, R., Vandenberghe, A., Merregaert, J., Contreras, R. and Fiers, W. (1971) Proc. Natl. Acad. Sci. U.S. 68, 585.
[8] De Wachter, R., Merregaert, J., Vandenberghe, A., Contreras, R. and Fiers, W. (1971) European J. Biochem. 22, 400.
[9] De Wachter, R. and Fiers, W. (1971) Methods in Enzymology, (Grossman L. and Moldave, K., eds.), Vol. 21, p. 167, Academic Press.
[10] De Wachter, R. and Fiers, W. (1972) Anal. Biochem. 49, 184.
[11] Contreras, R. and Fiers, W. (1971) FEBS Letters 16, 281.
[12] Sanger, F., Brownlee, G.G. and Barrell, B.G. (1965) J. Mol. Biol. 13, 373.
[13] Brownlee, G.G. and Sanger, F. (1967) J. Mol. Biol. 23, 337.
[14] Adams, J.M., Jeppesen, P.G.N., Sanger, F. and Barrell, B.G. (1969) Nature 223, 1009.

[15] Haegeman, G., Min Jou, W. and Fiers, W. (1971) J. Mol. Biol. 57, 597.
[16] Adams, J.M., Spahr, P.F. and Cory, S. (1972) Biochemistry 11, 976.
[17] Weiner, A.M., Platt, T. and Weber, K. (1972) J. Biol. Chem. 247, 3242.
[18] Berissi, H., Groner, Y. and Revel, M. (1971) Nature New Biol. 234, 44.
[19] Lodish, H.F. (1971) J. Mol. Biol. 56, 627.
[20] Argetsinger-Steitz, J. (1973) Proc. Natl. Acad. Sci. U.S., in press.
[21] Clark, B.F.C. and Marcker, K.A. (1966) J. Mol. Biol. 17, 394.
[22] Ghosh, H.P., Söll, D. and Khorana, H.G. (1967) J. Mol. Biol. 25, 275.
[23] Stewart, J.W., Sherman, F., Shipman, N.A. and Jackson. M. (1971) J. Biol. Chem. 246, 7429.