STUDIES ON THE BACTERIOPHAGE MS2. SOME NUCLEOTIDE SEQUENCES FROM THE RNA-POLYMERASE GENE*

R. CONTRERAS, A. VANDENBERGHE, G. VOLCKAERT, W. MIN JOU and W. FIERS

Laboratory of Molecular Biology and Laboratory of Physiological Chemistry University of Ghent, 9000 Ghent, Belgium

Received 16 June 1972

1. Introduction

The knowledge of the amino acid sequence from the coat protein of the RNA bacteriophages f2, R17 and MS2 [1–3] allowed the identification of several RNA fragments as being derived from this gene. They were all obtained by partial ribonuclease hydrolysis of R17 RNA [4–6], f2 RNA [7] or MS2 RNA [8, 9]. For MS2, these sequences could be further extended, and the complete sequence of the coat gene was established [10]. Furthermore, this sequence was followed by a 36 nucleotides long intercistronic region and by the ribosomal binding site of the next gene, the RNA-polymerase. As the amino acid sequences for the two other gene products, the A-protein and the viral RNA-polymerase, are not known, no such simple test exists to identify polynucleotides corresponding to the latter two genes.

In this communication we present data on three nucleotide fragments derived from the polymerase gene. This conclusion is based on direct overlaps with known sequences [10], on the established gene order for these phages [11, 12], and on the mapping of a series of polypurine tracts in two fragments of the RNA molecule [13].

2. Methods

All RNA fragments described in this study were obtained from partial ribonuclease T1 hydrolysates of ^{32}P-labelled MS2 RNA [14]. They were purified either by electrophoresis on acidic polyacrylamide gel in the presence of urea [15] or by two-dimensional gel electrophoresis [16]. The structure determination of the pure fragments will be published elsewhere and was done mainly according to the methods developed by Sanger and co-workers [4, 17, 18]. The relative ordering of the oligonucleotides in one of the fragments (B9z1) was determined by partial digestion with CM-RNAase* and has already been reported [19].

3. Results and discussion

Extensive characterization of many purified RNA fragments (see Methods section) led to the construction of three RNA segments with chain lengths of respectively 105 (fig. 1), 79 (fig. 2) and 112 nucleotides (fig. 3). Splitting points for T1-ribonuclease observed under different partial digestion conditions are indicated by arrows.

The sequence in fig. 1 is directly identified as the beginning of the polymerase gene as it contains part of the ribosome binding site of this gene (between arrows 0 and 2) [10, 20]. Moreover, this sequence was also found as a part of the fragment β4b1 (described in [10]); this is a polynucleotide of around 180 nucleotides in length, which starts inside the

* Part XIX of a series.

* CM-RNAase is ε-carboxymethyllysine-41-pancreatic ribonuclease A.

North-Holland Publishing Company – Amsterdam

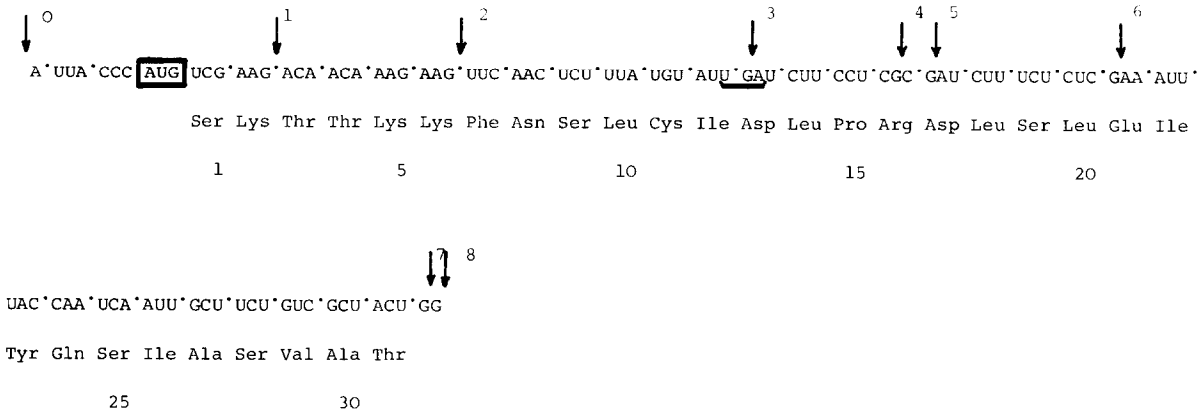

Fig. 1. Beginning of the polymerase gene. Arrows (with corresponding numbers) point to different splitting points under our partial digestion conditions. Seven different fragments from this region have been isolated.

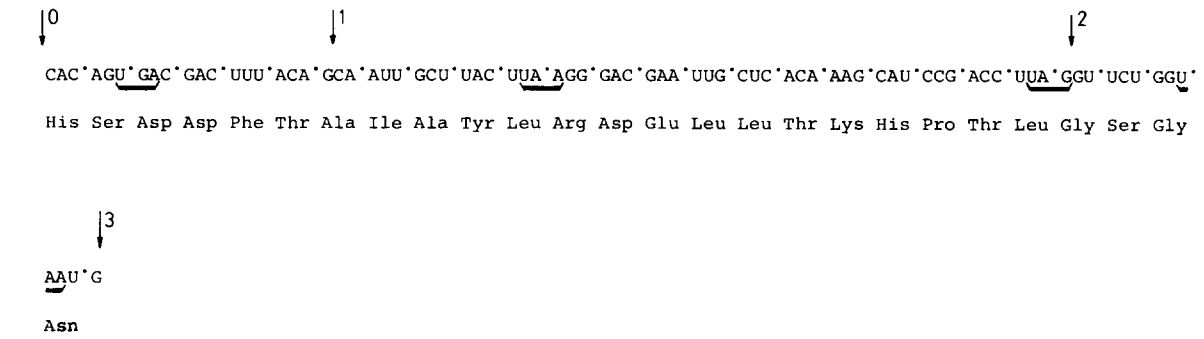

Fig. 2. B9z1 region of the polymerase gene. Three different fragments were used for the construction of this segment.

coat gene and extends even a few nucleotides to the right of the sequence presented in fig. 1.

The segment B9z1 (fig. 2) has recently been found in longer RNA fragments which contained the polypurine tract AGAGAAGGU. Since it is known that this sequence is derived from the 21 S (3′-terminal) part of the MS2 RNA molecule [13], which comprises most of the coat gene together with the gene for the polymerase [11], the B9zl sequence must be part of the polymerase gene.

The γ4b7 segment (fig. 3) is derived from a region between 500 and 800 nucleotides from the 3′-end of the RNA chain. This conclusion is based on analysis of long polynucleotides obtained by very limited ribonuclease T1 hydrolysis of the RNA (to be described in detail elsewhere). Since the polymerase gene is known to be proximal to the 3′-end of the RNA, and since the limited genetic information excludes the possibility of an untranslated sequence of over 500 nucleotides at this 3′-end, these observations locate this γ4b7 segment in the polymerase gene (according to these studies, the B9z1 segment would be in the first half of the gene).

The choice of the reading frame for the beginning of the gene is obvious, as it contains the AUG initiator signal. For the two other sequences there is only one possible reading frame, as both other frames are blocked by one or several nonsense codons (cf. figs. 1, 2 and 3). The different codons used in the parts of the polymerase gene here described are listed in table 1. Together with the codons from the coat gene [10], all triplets have been

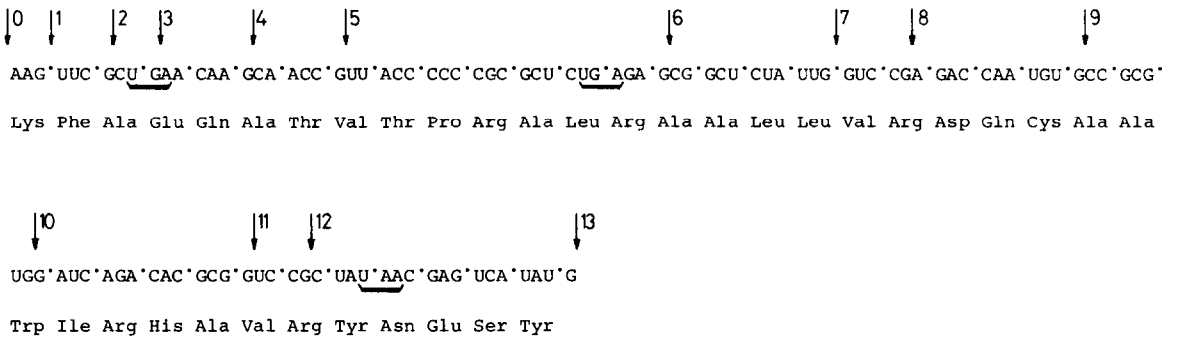

Fig. 3. γ4b7 segment of the polymerase gene. Sixteen different fragments have been isolated from this part of the molecule.

Table 1

CODONS FOUND IN POLYNUCLEOTIDE FRAGMENTS DERIVED FROM THE POLYMERASE CISTRON

	U	C	A	G	
U	Phe {o, oo} Leu {ooo, ●●}	Ser {oooo, oo, o}	Tyr {●●, oo} Ochre Amber	Cys {oo} Opal Trp o	U C A G
C	Leu {oo, oo, o, ●}	Pro {o, o, o}	His {●, ●●} Gln {ooo}	Arg {ooo, ●}	U C A G
A	Ile {oooo, o} Met ◎	Thr {o, ooo, ●●●●}	Asn {o, oo} Lys {ooooo}	Ser {●} Arg {●●, ●}	U C A G
G	Val {o, ooo}	Ala {88oo, o, oo, ooo}	Asp {oo, oooo} Glu {ooo, o}	Gly {oo}	U C A G

Filled symbols denote codons not used in the coat gene.

used at least once, except AUA (Ile) and CGG (Arg). There are indications that *E. coli* tRNA can indeed not respond to the AUA codon for isoleucine [21].

Several codons *not* used to specify the coat gene are present in the polymerase gene: UUG (Leu), CUG (Leu), ACA (Thr), UAU (Tyr), CUU and CUC (His), CGA (Arg), AGU (Ser), AGA and AGG (Arg). For the histidine codons this is merely a reflection of the different amino acid composition of the proteins, and for still others it may be due to a coincidence. But, if a modulation-type control [22] exists (controlling the amounts of different proteins by the appropriate choice of codons used to specify them) candidates as rate-limiting codons are most likely to be present in the above series (in infected cells much less viral polymerase polypeptide is made than coat protein). There is indeed evidence for the arginine codons AGA and AGG that the rate of translation is limited by the concentration of corresponding tRNA [23].

Acknowledgement

The research project was supported by the Fonds voor Kollektief Fundamenteel Onderzoek of Belgium (no. 841). R.C. and G.V. thank the NFWO and the IWONL respectively for fellowships.

References

[1] K. Weber and W. Konigsberg, J. Biol. Chem. 242 (1967) 3563.
[2] K. Weber, Biochemistry 6 (1967) 3144.
[3] J. Vandekerckhove and M. van Montagu, in preparation.
[4] J.M. Adams, P.G.N. Jeppesen, F. Sanger and B.G. Barrell Nature 223 (1969) 1009.
[5] J.L. Nichols, Nature 225 (1970) 147.
[6] P.G.N. Jeppesen, J.L. Nichols, F. Sanger and B.G. Barrell, Cold Spring Harbor Symp. Quant. Biol. 35 (1970) 13.

[7] J.L. Nichols and H.D. Robertson, Biochim. Biophys. Acta 228 (1971) 676.
[8] W. Min Jou, G. Haegeman and W. Fiers, FEBS Letters 13 (1971) 105.
[9] W. Fiers, R. Contreras, R. de Wachter, G. Haegeman, J. Merregaert, W. Min Jou and A. Vandenberghe, Biochimie 53 (1971) 495.
[10] W. Min Jou, G. Haegeman, M. Ysebaert and W. Fiers, Nature 237 (1972) 82.
[11] P.G.N. Jeppesen, J. Argetsinger Steitz, R.F. Gesteland and P.F. Spahr, Nature 230 (1970) 1.
[12] R.N.H. Konings, R. Ward, B. Francke and P.H. Hofschneider, Nature 226 (1970) 604.
[13] W. Min Jou, W. Fiers, H. Goodman and P. Spahr, J. Mol. Biol. 42 (1969) 143.
[14] R. de Wachter, J. Merregaert, A. Vandenberghe, R. Contreras and W. Fiers, European J. Biochem. 22 (1971) 400.
[15] R. de Wachter and W. Fiers, in: Methods in Enzymology, Vol. 21, Part D (1971) p. 167.
[16] R. de Wachter and W. Fiers, Anal. Biochem., in press.
[17] F. Sanger, G.G. Brownlee and B.G. Barrell, J. Mol. Biol. 13 (1965) 373.
[18] G.G. Brownlee and F. Sanger, J. Mol. Biol. 23 (1967) 337.
[19] R. Contreras and W. Fiers, FEBS Letters 16 (1971) 281.
[20] J. Argetsinger Steitz, Nature 224 (1969) 957.
[21] C.T. Caskey, A. Beaudet and M. Nirenberg, J. Mol. Biol. 37 (1968) 99.
[22] B.N. Ames and P.E. Hartman, Cold Spring Harbor Symp. Quant. Biol. 35 (1963) 349.
[23] W.F. Anderson, Proc. Natl. Acad. Sci. U.S. 62 (1969) 566.