STUDIES ON THE BACTERIOPHAGE MS2. NUCLEOTIDE FRAGMENTS FROM THE COAT PROTEIN CISTRON*

W. MIN JOU, G. HAEGEMAN, and W. FIERS
Laboratory of Molecular Biology and Laboratory of Physiological Chemistry, University of Ghent, Belgium

Received 21 December 1970

1. Introduction

Partial hydrolysis with ribonuclease T1 has already been used for the isolation of fragments derived from different regions of R17 [1–4] or of MS2 RNA [5,6]. In this paper, we describe the isolation and purification of several RNA fragments coding for different parts of the amino acid sequence of the coat protein. The structural data obtained upon sequence analysis of the purified fragments (not presented here) are compared with the nucleotide sequences from the corresponding parts of R17 RNA [1–3]. The polynucleotides studied have a hairpin-like structure with considerable base pairing, as was found for all the RNA segments isolated from partial digests until now [1–6].

The results further support the concept that the evolutionary process has taken advantage of the degeneracy of the genetic code in order to maximize secondary structure in this type of messenger RNA, as proposed originally in reference [1].

2. Methods

^{32}P-labelled MS2 RNA was partially degraded with ribonuclease T1 and the hydrolysate separated on a 12% polyacrylamide gel slab [7]. RNA from different gel bands was hydrolyzed to completion with ribonuclease T1 and fingerprinted, using electrophoresis at pH 3.5 in one dimension, followed by chromatography on DEAE-paper (with 'homomixture c') as the second [1]. A similar mapping system was used for the purification of RNA fragments. For that purpose, the second dimension was on thin-layer plates using homomixture a [8]. An alternative method for further purification consists of gel-electrophoresis at pH 3.5 [7]. Conditions for both fractionation systems are given in the legend of fig. 1.

For the sake of completeness, the methods used for the structure determination will be briefly outlined here, but these results will be described in detail elsewhere. The digestion conditions and separation systems used were as described by Sanger and coworkers [1, 9, 10].

(a) A ribonuclease T1 map of the fragment was prepared. The RNase T1 oligonucleotides were studied further by pancreatic and by RNase U2 degradation, and by modification with a carbodiimide reagent (*N*-cyclohexyl-*N'*-(β-morpholinyl-(4)-ethyl) carbodiimide-methyl-*p*-toluene sulphonate) followed by pancreatic RNase cleavage. If necessary, partial degradations with venom and/or spleen exonuclease were undertaken.

b) A pancreatic ribonuclease map was prepared, and the products were characterized with RNase T1 (and spleen exonuclease, if necessary).

c) The RNA sample was partially degraded with

* Part XI

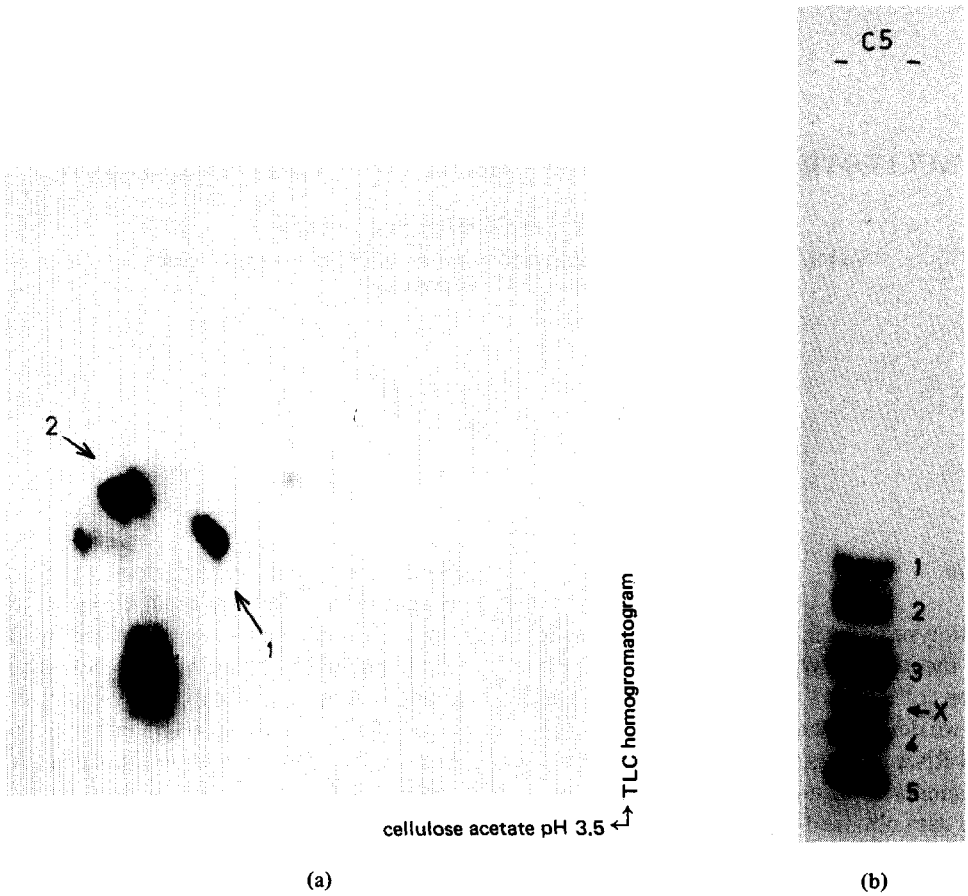

Fig. 1. (a) Two-dimensional fractionation of RNA fragments from band C5. Separation in the first dimension was by electrophoresis on cellulose acetate at pH 3.5 in 7 M urea. Material in the region between the blue and pink marker was transferred on to a thin-layer plate, which consisted of a mixture of DEAE-cellulose and cellulose in a ratio of 1:10. The solvent in the second dimension was a 5 percent RNA solution in 7 M urea (homomixture a, [8]). 1 and 2 show the positions of the fragments corresponding to the coat protein cistron. b) Polyacrylamide gel electrophoresis under acidic conditions. 20 μl, containing the RNA extracted from band C5, 400 μg carrier yeast RNA, 7 M urea, sucrose and xylene cyanol FF, was loaded in 2 X 20 mm slots on a 10% polyacrylamide gel. The buffer contained 0.025 M citric acid and 6 M urea (both in the gel and in the reservoirs). Electrophoresis was for 18 hr at 600 V. 1 marks the position of the intact fragment. 3 and 5 are the two halves of the RNA hairpin. X shows the position of the xylene cyanol marker.

ribonuclease T1 in the following conditions: 50 to 80 μg of RNA and between 0.2 and 1.0 unit of enzyme were incubated in 5 μl 0.1 M sodium phosphate buffer, pH 7.4 + 0.002 M EDTA for 15–30 min at 0°. The partial digest was separated on a mapping system using 'homochromatography' on paper or on a thin-layer plate (described above) with homomixtures a, b or c of reference [8] as eluents. The partial RNase T1 products were characterized by degradation with T1 and pancreatic ribonuclease.

The partial T1 fragments (section c) together with the pancreatic ribonuclease products (section b) provided sufficient information to reconstruct the sequences here described.

3. Results

Since the amino acid sequence of the viral coat protein cistron is known [11–13], one can write up a set of alternative nucleotide sequences for this cistron by applying the genetic code. On this basis, we could then search for oligonucleotides related to this phage protein in different bands of the partial digest.

The longest oligonucleotides present in the electrophoretic-homochromatographic fingerprint of a ribonuclease T1 digest of each band (this kind of separation gives a maximal resolution of the longer G-terminal oligonucleotides; even from relatively complex mixtures) were sequenced (see Methods) until we could decide if they did (or did not) fit the coat protein cistron. A similar approach has been used for the isolation of fragments derived from the coat protein cistron of R17 RNA [1–3].

In this way, oligonucleotides coding for the following amino acid sequences from the coat protein were found (the numbers refer to the amino acids in the coat polypeptide): 1–6 (band D6), 34–37 and 41–43 (band C5), 56–61 and 68–70 (band C6), 85–88 and 89–95 (band C7), 124–126 and 127–129, followed by the chain termination signal (band D1). The last band (D1) is indicated in fig. 3a of reference [7] which shows a typical polyacrylamide gel pattern of a partial digest of Ms2 RNA. C6, C6, and C7 immediately precede it, while D6 is the band near the bromophenol blue marker.

As judged from the position on the gel, band C5 should be around 60 and D6 around 30 nucleotides long. The RNase T1 fingerprints, however, indicated considerably more complexity, which meant that we were still dealing with mixtures, and further purification was necessary (see Methods section).

Separations of the ^{32}P-material from band C5 by both methods are shown in fig. 1. Spot no.1 from the thin layer plate as well as band 3 from the gel contained the two 'reference' oligonucleotides found in C5 and are in fact identical. The structure of this 34-long fragment (determined as outlined in Methods) corresponds with the amino acids 33 to 43. Spot no.2 (or band 5 of the gel) was found to code for the amino acid residues 44 to 52, directly adjacent to the former fragment. In this way, it was possible to construct the RNA sequence coding for amino acids 33–52. It can be written as a hydrogen-bonded hairpin (fig. 2). This whole sequence was also present in an intact, albeit very impure form, in band 1 of the gel. We believe that most of this fragment is present on the neutral gel as a nicked hairpin-like structure held together by hydrogen bonds. Both strands are separated on rerunning, since either of the purification systems involve denaturing conditions (pH 3.5 and 7 M urea).

A similar situation was encountered for band C6. Upon further purification, using the 'mapping' technique, two fragments derived from the coat protein cistron were found. Placed contiguously, they constitute a 59 nucleotide long sequence which corresponds to amino acids 57 to 75. It can again be arranged as a hairpin-like structure (fig. 2). No trace of the intact loop (to be expected in the slower moving material on the plate) was detected.

A 57- (coding for amino acids 81-99) and a 43-long fragment (amino acids 124–129) were purified by electrophoresis on acidic polyacrylamide gels from bands C7 and D1 respectively, and their structure determined. Another oligonucleotide, corresponding to amino acids 1 to 6, was present in the digest of band D6. It has so far not been isolated from a purified, longer treatment.

The sequences of the polynucleotides, whose purification is described in this paper, are summarized in fig. 2.

4. Discussion

The nucleotide sequences presented here account for more than 50% of the coat protein cistron. All four RNA fragments can be arranged in hairpin-like structures with a considerable degree of base pairing. It is known from the work of Argetsinger-Steitz [14] that the beginning of the coat cistron in R17 RNA forms also a part of a looped structure. The major unknown parts of the cistron (on either side of the three central loops) could, according to their size, give rise to similar hairpin structures.

Whatever the need for this high degree of secondary structure may be, one can wonder how this is realised in a messenger RNA, which specifies a defined amino acid sequence. It has been suggested that this is one of the functions of the genetic code [1], by appropriate choice of the third letters. In fact, the

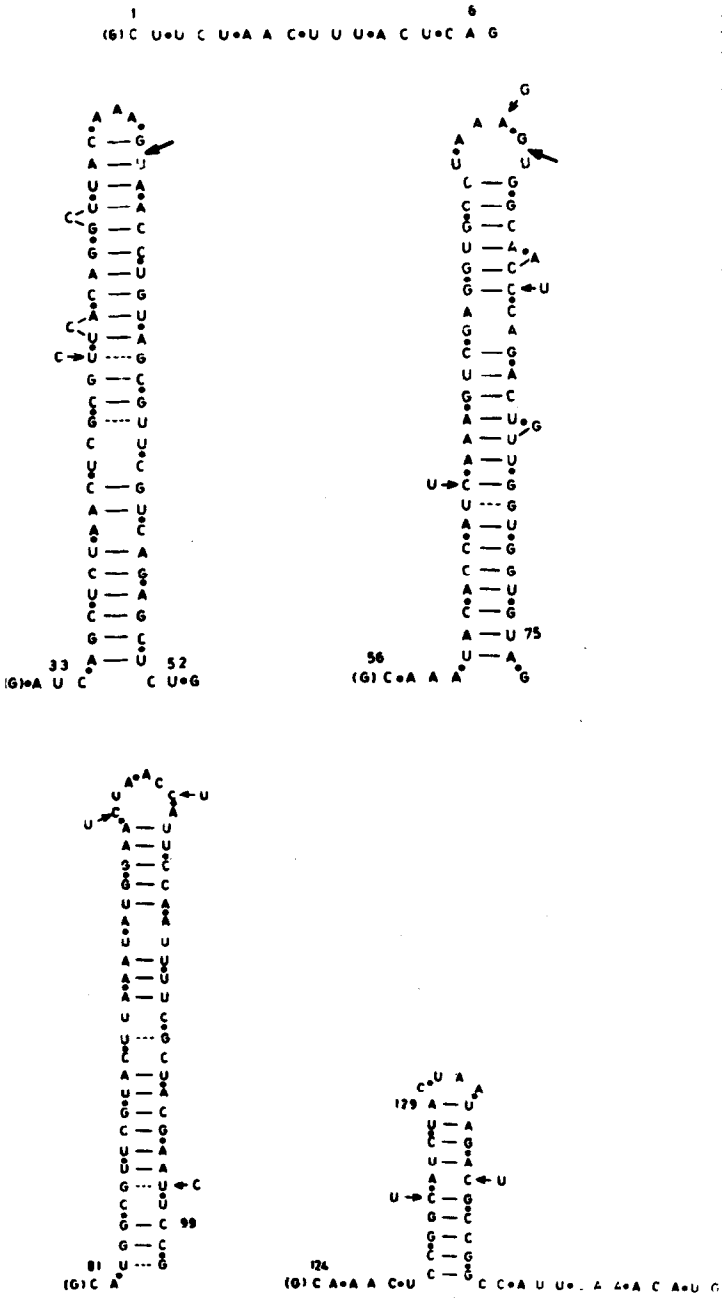

Fig. 2. Structure of RNA sequences corresponding to different parts of the MS2 coat protein cistron. The numbers refer to the amino acids in the coat polypeptide. Nine nucleotides, which are different in R17 RNA, are indicated by thin arrows. The nucleotides found in the latter phage are given beside. Heavy arrows at the top of the first two hairpins show the position of the nick in these structures.

percentage of third codon letters involved in base pairing (77.6%) is definitely higher than the average (70.0%) for the three internal loops of the cistron known until now. Third letters of codons are only rarely found opposite to one another in these structures, in only 5 out of 74 base pairs are both partners third letters (stastically one would expect one third). This makes it possible to change (almost) two-thirds. of the base pairs (by a mutation in the third letter participating in it) with no effect on the amino acid sequence specified. The comparison of the coat protein sequences from MS2 RNA with the corresponding regions from R17 [1–3] reveals only nine single base changes (fig. 2). Seven are third letter changes, one codon differs in the first letter (a leucine codon at position 90: CUA ↔ UUA), and one nucleotide change is found in the region following the chain termination signal of the cistron. In the latter case, the sequence UAA·UAG·AUG found in R17 RNA is changed into UAA·UAG·ACG, thus abolishing the possible new chain initiator AUG.f2 RNA has the same sequence as MS2 in this region (J.L. Nichols and H.D. Robertson, unpublished observation cited in [3]). Therefore, the hypothetical hexapeptide coded for by the intercistronic region between the coat protein and the synthetase in R17 cannot be made in MS2 (and in f2). The chain termination signal of the coat protein cistron consists of the same two successive nonsense codons, as found both in R17 and in f2 RNA.

The nine base changes are all transitions, eight of which being pyrimidine–pyrimidine interchanges. There is a net loss of 3 A–U pairs in MS2 as compared with R17, respectively one in the regions corresponding to amino acids 57–75, 81–99 and in the chain termination region (the structure of the RNA fragment coding for amino acids 57–75 is shown in a slightly altered base pairing scheme as compared with the corresponding R17 RNA region).

It is obvious that the hereditary phage message has been written up in a highly degenerate code, but more information about this (and other) cistron(s) will be needed in order to allow conclusions concerning the frequency of occurrence of different codons.

Acknowledgements

G.H. and W.M. hold a fellowship from the Nationaal Fonds voor Wetenschappelijk Onderzoek (Belgium). The research was supported by a grant from the Fonds voor Kollektief Fundamenteel Onderzoek (no 841). We thank Mr. F. Duerinck, Mr. A. Raeymaekers and Mrs. M. Borremans for excellent technical assistance.

References

[1] J.M. Adams, P.G.N. Jeppesen, F. Sanger and B.G. Barrell, Nature 223 (1969) 1009.
[2] J.L. Nichols, Nature 225 (1970) 147.
[3] P.G.N. Jeppesen, J.L. Nichols, F. Sanger and B.G. Barrell, Cold Spring Harbor Symp. Quant. Biol. (1970) in press.
[4] J.M. Adams and S. Cory, Nature 227 (1970) 570.
[5] W. Min Jou, R. Contreras and W. Fiers, FEBS Letters 9 (1970) 222.
[6] R. De Wachter, A. Vandenberghe, J. Merregaert, R. Contreras and W. Fiers, in press.
[7] R. De Wachter and W. Fiers, Methods in Enzymology, Vol. 21, in press.
[8] G.G. Brownlee and F. Sanger, European J. Biochem. 11 (1969) 395.
[9] F. Sanger, G.G. Brownlee and B.G. Barrell, J. Mol. Biol. 13 (1965) 337.
[10] G.G. Brownlee and F. Sanger, J. Mol. Biol. 23 (1967) 337.
[11] K. Weber, Biochemistry 6 (1967) 3144.
[12] J.Y. Lin, C.M. Tsung and H. Fraenkel-Conrat, J. Mol. Biol. 24 (1967) 1.
[13] J. Vandekerckhove, H. Francq and M. Van Montagu, Arch. Intern. Physiol. Biochim. 77 (1969) 175.
[14] J.A. Steitz, Nature 224 (1969) 957.