The Leader Sequence from the 5'-Terminus to the A-Protein Initiation Codon in MS2-Virus RNA

R. DE WACHTER, A. VANDENBERGHE, J. MERREGAERT, R. CONTRERAS, AND W. FIERS

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

Communicated by Robert L. Sinsheimer, December 21, 1970

ABSTRACT RNA fragments of different chain length, each containing the 5'-terminal guanosine tetraphosphate (pppGp) of bacteriophage-MS2 RNA, have been isolated from partial ribonuclease digests of the viral RNA. The longest fragment overlaps with the ribosomal-binding site of the A-protein cistron. The base sequence has been established for the major part. The results directly confirm that the A-protein cistron is closest to the 5'-terminus. Its initiating (AUG) codon starts at position 130, being preceded by an untranslated sequence of 129 nucleotides.

Considerable progress has been made in the sequence determination of bacteriophage RNA molecules by partial ribonuclease hydrolysis and fractionation of the fragments by polyacrylamide gel electrophoresis. The possibility of obtaining large and specific fragments from RNA molecules by this method was demonstrated on an analytical scale by McPhie, Hounsell, and Gratzer (1) and by Gould (2) for ribosomal RNA and by Min Jou, Hindley, and Fiers (3) for bacteriophage RNA. Applying the technique on a preparative scale to R17 RNA, Adams et al. (4) and Nichols (5) determined the sequence of a part of the coat-protein cistron and of its termination region, and Adams and Cory (6) reported the sequence extending from the 5'-terminus to position 74 of the RNA chain. By comparison with the ribosomal-binding regions that had been elegantly isolated and sequenced by Steitz (7), Adams and Cory could deduce that the initiation codon for the first cistron, the A-protein (8), is preceded by an untranslated sequence of at least 91 nucleotides. Similarly, the 5'-terminal sequence of Qβ RNA determined by Billeter et al. (9), who used partial in vitro synthesis, is untranslated over a distance of at least 61 nucleotides.

We have now isolated 5'-terminal fragments, of chain lengths up to 125, from partial hydrolysates with T1 ribonuclease of bacteriophage-MS2 RNA. The sequence is identical with that of R17 RNA as far as R17 has been reported (6) and, moreover, it overlaps at the 3'-end with the sequence of the A-protein initiation region of R17 RNA (7,8). This allows us to deduce that the untranslated 5'-terminal sequences of both phages are largely, if not completely, identical and have a length of 129 nucleotides.

METHODS

32P-labeled bacteriophage MS2 and its RNA were prepared according to Fiers, Lepoutre, and Vandendriessche (10). Partial digests of RNA by T1 ribonuclease were separated by electrophoresis on polyacrylamide-gel slabs at pH 8, and RNA fragments were purified on gel slabs containing 7 M urea at pH 3.5. Preparation of gels, electrophoresis, autoradiography, and recovery of RNA from the gel were described by De Wachter and Fiers (11).

![Fig. 1. Isolation of 5'-terminal fragments from MS2 RNA.](image-url)
Complete T1 ribonuclease digests of isolated RNA fragments were separated by two-dimensional fingerprinting according to Sanger and Brownlee (12); the first dimension consisted of electrophoresis on cellulose acetate at pH 3.5 in the presence of 7 M urea and the second of electrophoresis on DEAE-cellulose paper in 7% formic acid. For pancreatic ribonuclease digests, the first dimension and the blotting procedure were the same, but the second dimension consisted of descending chromatography with a 90-ml concentration gradient (13) from 0.05 M Tris-formate, 0.025 M Tris, 8 M formamide to 0.45 M Tris-formate, 0.225 M Tris, 8 M formamide.

The sequence of the resulting oligonucleotides was determined by T1, or pancreatic ribonuclease hydrolysis, partial hydrolysis with spleen phosphodiesterase (12), U2 ribonuclease hydrolysis, and carbodiimide-blocked pancreatic ribonuclease hydrolysis (4).

RESULTS

Isolation of 5'-terminal fragments of MS2 RNA

Hydrolysis of 32P-labeled MS2 RNA with T1 ribonuclease at 0°C yields a mixture of partial hydrolysis products, which can be fractionated roughly according to molecular weight by electrophoresis at pH 8 on a polycrylamide-gel slab. An autoradiogram of such a gel is shown in Fig. 1a. When the material present in a particular band is extracted and subjected to a second electrophoresis at pH 3.5 in a gel containing 6 M urea, it is further separated into a number of individual components, which are now sufficiently homogeneous for sequence analysis. Four bands were found to contain fragments carrying the 5'-terminal pppGp of MS2 RNA. These bands are indicated in Fig. 1a, and their purification on the acid gel is shown in Fig. 1b. The 5'-terminal fragments were identified by assaying the bands for their pppGp content after alkaline hydrolysis (13) and by comparing the oligonucleotide patterns obtained after T1 ribonuclease hydrolysis and two-dimensional fingerprinting. After sequence analysis, the fragments were found to have chain lengths of 125, 121, 82, 74, and 57 nucleotides, respectively.

Sequence determination of the isolated fragments

A pancreatic ribonuclease fingerprint of the fragment of chain-length 125 is shown in Fig. 2 and a T1 ribonuclease fingerprint in Fig. 3. Similar fingerprints were prepared for the smaller fragments and the sequence of the constituent oligonucleotides was determined.

The relative ordering of the oligonucleotides in the fragments was deduced mainly by partial hydrolysis with T1 ribonuclease. The resulting products were again separated on a polyacrylamide gel, an autoradiograph of which is shown in Fig. 4. The oligonucleotide composition of each partial-digest product was determined by T1 ribonuclease digestion followed by two-dimensional fingerprinting, and by pancreatic ribonuclease digestion followed by one-dimensional electrophoresis on DEAE-cellulose paper at pH 1.9.

On the basis of these results, the partial nucleotide sequence shown in Fig. 5 could be reconstructed.
Comparison of the 5'-terminal sequences of MS2 and R17 RNA

The 5'-terminal sequence of MS2 RNA is probably entirely identical with the corresponding sequence of R17 RNA as published up to position 74 (b). Some uncertainty remains in two short pyrimidine stretches in MS2 RNA but, as their base composition is the same as in the case of R17 RNA, they will very likely turn out to be identical. We therefore expect both sequences to remain very similar, if not identical, further away from the terminus, towards the initiation site of the first cistron. In the case of R17 RNA and M12 RNA, the order of the genes has been determined as 5'-terminus-A protein-coat protein—polymerase-3'-terminus (8,14), and the base sequence of the ribosomal-attachment site for the R17 RNA A-protein cistron is known (7,8).

Comparison of the R17 attachment site with the 5'-terminal fragment of MS2 RNA reveals that positions 110 to 125 of MS2 RNA overlap with the first 16 bases of the R17 RNA initiation site. Such a long overlap is very unlikely to occur by chance, especially as the heptanucleotide AGGAGGU occurs only once in the entire MS2 RNA sequence (15). We consider that the existence of this overlap, together with the identity of the 74-base fragments of MS2 and R17 RNA, provide a sufficient argument for locating the position of the A-protein initiation site of both bacteriophages in the RNA chains. Letting the corresponding parts of the MS2 and R17 sequences coincide (Fig. 5), we obtain position 130 for the initiation codon, AUG, of the A-protein cistron, which establishes the chain length of the untranslated 5'-terminal sequence at 129. The overlapping of the MS2 5'-terminal and
Biochemistry: De Wachter et al.

The R17 initiation site sequences directly confirms that the A-protein cistron is located closest to the 5'-terminus of the three genes present in the RNA chain.

DISCUSSION

Adams and Cory (6) have pointed out the analogies that exist between the 5'-terminal sequence of R17 and Q8 RNA, as well as the internal homologies in the two sequences, and have postulated that both phages could have evolved from a common ancestor. Even more striking is the identity between the R17 and MS2 RNA 5'-terminal sequences. Indeed, not a single base difference occurs over a length of 74 residues, whereas in the coat-protein cistron, seven base differences are noted of a total of 194 residues so far sequenced (ref. 16 and W. Min Jou and G. Haegeman, personal communication). This is noteworthy because the coat protein cistrons code for identical amino acid sequences. On the other hand, in the untranslated leader stretch, the selection pressure to assure synthesis of a fixed amino acid sequence does not exist; nevertheless, the primary structure is entirely unaffected by mutations. It could be that the rigorous conservation of the sequence of this untranslated RNA is necessary in order to form some very specific tertiary structure at the 3'-terminus of the complementary minus strand, which could serve as a recognition site for the viral RNA polymerase.

A puzzling peculiarity seems worth mentioning, namely, that the distance in the leader portion between two termination codons, or between a termination codon and an AUG triplet, is frequently a multiple of three (Fig. 6). The same is true for the corresponding leader sequence of Q8 RNA (9). The meaning of initiation and termination codons, and of triplet codons, are all defined as a function of their role in protein synthesis, and their significance, if any, in the context of an untranslated nucleotide sequence is unknown.

Fig. 4. Partial ribonuclease $T_1$-hydrolysis of 5'-terminal fragments. The fragments of chain length 125, 121, 74, and 37, which had an activity of 1-5 x $10^6$ dpm, were digested with 40 units of $T_1$-ribonuclease in the presence of 0.2 mg of carrier yeast RNA for 20 min at 20°C, in 0.03 ml of 0.1 M Tris-HCl (pH 7.5) + 0.1 mM EDTA. After phenol extraction, the digests were separated by electrophoresis overnight at 500 V on a 20% polyacrylamide-gel slab containing 0.04 M Tris-acetic acid (pH 8).

The composition of partial products T1 to T7 can be found in Fig. 5. It is apparent from the autoradiograph that the different fragments give a related pattern of partial products, and that the smaller fragments originate by partial digestion of the larger ones, e.g., partial-product T2 obtained from the 74-base fragment is identical with the 37-base fragment. The bands running faster than T7 are complete hydrolysis products containing one G residue. The only difference between the fragments of chain-length 125 and 121 is the presence of UUUG in the former.

The MS 2...UAA...3x3...UAA...2x3...UAG...4x3...AUG

...UAG...2x3...UGA...1x3...AUG

Q8...UGA...1x3...UAA

...UAA...10x3...UAAUGA...2x3...UAAUGA

Fig. 6. Distances, expressed in number of nucleotides, between (nonfunctional ?) termination and initiation codons present in the untranslated 5'-leader sequence of MS2 RNA and Q8 RNA. The distances are often a multiple of three. Not considered are two UAG triplets in MS2 RNA (because their exact location is unknown), an UAG and an AUG triplet in Q8 RNA (position 14 and 62, respectively), and AUG triplets that overlap with UAA triplets.

This research was supported by grants from the Fonds voor Kollektief Fundamenteel Onderzoek (no. 841) and from the U.S. National Institutes of Health (GM 11304). We thank Mr. R. De Baere, A. Raeymaekers, L. Van Puyvelde, and Mrs. M. Borremans-Bensch, who participated in various parts of this work.
FIG. 5. The 5'-terminal sequence of MS2 RNA. The arrows indicate the end of the different 5'-terminal fragments that were isolated, and the corresponding numbers, the bands from which they were purified (Fig. 1). Products T1 to T7 were obtained by partial T1-ribonuclease hydrolysis of the fragments (Fig. 4). The oligonucleotides at the 3'-end of the sequence were ordered on the basis of partial pancreatic-ribonuclease hydrolysis. Mono- or oligonucleotides separated by commas may have to be permuted within the brackets to obtain the actual sequence. The R17 RNA A-protein initiation region (7,8) has been written under the MS2 RNA sequence so as to show the overlapping part.