Proc. Nat. Acad. Sci. USA
Vol. 72, No. 11, pp. 4252–4256, November 1975
Biochemistry

Nucleotide sequence of microvariant RNA: Another small replicating molecule

(Qβ replicase/endogenous reaction product/duplex blocks/logical choice sets/fragment length map)

D. R. MILLS, F. R. KRAMER, C. DOBKIN, T. NISHIHARA, AND S. SPIEGELMAN

Institute of Cancer Research and Department of Human Genetics and Development, College of Physicians & Surgeons, Columbia University, New York, N.Y. 10032

Contributed by S. Spiegelman, August 1, 1975

ABSTRACT **Microvariant RNA, a small self-replicating molecule (114 nucleotides long), has been isolated from Qβ replicase reactions incubated in the absence of exogenous template. Its complete nucleotide sequence has been determined. Comparison with MDV-1 RNA, a somewhat larger endogenous Qβ replicase product (220 nucleotides long) that had previously been characterized, revealed no significant sequence similarity. Since Qβ replicase can mediate the synthesis of both of these disparate RNA molecules, primary sequence cannot be the sole determining factor in the processes of enzyme recognition and replication. This implies that the key is to be found in the secondary or tertiary structures. The availability of two different replicating molecules of defined sequence should aid in identifying these critical structural features.**

Qβ replicase is a template-specific RNA-directed RNA polymerase induced in *Escherichia coli* infected with the RNA bacteriophage, Qβ. Its successful isolation (1) made possible the *in vitro* autocatalytic synthesis (2) of biologically competent viral RNA (3) in which the genetic type synthesized was determined by the initiating template RNA (4). This system permitted the performance of Darwinian selection experiments with replicating molecules (5). By imposing different selective pressures, variants emerged possessing a variety of heritable differences (6–8).

When Qβ replicase is incubated in the absence of exogenous template RNA, an endogenous product (MDV-1 RNA) can be isolated from the reaction (9). Its small size (220 nucleotides) led to the complete determination of its nucleotide sequence (10). The extensive occurrence in this molecule of secondary structures formed from intrastrand complements made it an ideal subject for studying the relationship between RNA structure and replicative function. Accordingly, MDV-1 RNA was used as the initiating template for selection experiments in which an altered function (resistance to ethidium inhibition of replication) was identified with particular base substitutions in its nucleotide sequence (11).

In the course of examining endogenous Qβ replicase reactions that gave rise to MDV-1 RNA, a smaller and apparently unrelated variant was also noted. The advantages inherent in comparing two divergent replicating RNA molecules impelled us to isolate and characterize this smaller (114 nucleotides), "microvariant" RNA. The present paper reports its complete sequence obtained by combining the classical two-dimensional fingerprinting methods (12–14) with the newer stratagems of nearest-neighbor analysis and "fragment length mapping" (10).

MATERIALS AND METHODS

Materials. Qβ replicase was isolated from Qβ-infected *E. coli* Q13 (14). Ribonucleoside [α-^{32}P]triphosphates, at specific activities greater than 50 Ci/mmol, were obtained from the International Chemical and Nuclear Corp., Irvine, Calif.

Reaction Conditions. Microvariant RNA was synthesized in the presence of: 84 mM Tris·HCl (pH 7.4), 12 mM $MgCl_2$, 200 μM (each) ribonucleoside triphosphate (one of which was isotopically labeled), 40–400 ng/ml of microvariant (+) RNA, and 64 μg/ml of Qβ replicase. Reaction mixtures were incubated at 37° for 10 min. The synthesized microvariant RNA consisted of complementary (+) and (−) strands, with an excess of (+) strands.

Strand Separation. Isolated microvariant RNA was self-annealed at 65° for 60 min and then separated by polyacrylamide gel electrophoresis into two species: double-stranded (+/−) RNA and single-stranded (+) RNA. The details of

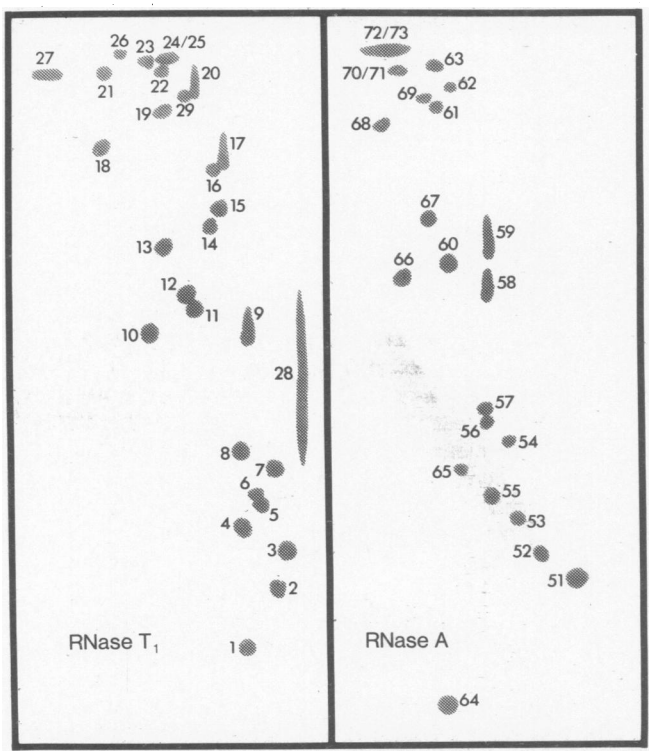

FIG. 1. Oligonucleotide fingerprint patterns of microvariant (+/−) RNA. Since each RNA preparation was labeled with only one of the four ribonucleosides and was then analyzed separately, no one autoradiograph would show the position of all the oligonucleotides. This figure is a composite tracing of separate fingerprints made with each of the four labels. Oligonucleotides obtained from RNase T_1 digests are numbered 1 through 29, and oligonucleotides obtained from RNase A digests are numbered 51 through 73.

RNA isolation, annealing, and electrophoresis have been described (11). For sequence analysis, (+/−) RNA was converted into an equimolar mixture of single-stranded (+) and (−) RNA by melting at 100° for 60 sec, followed by rapid freezing. Minus (−) RNA was isolated from this mixture by slab gel electrophoresis (15) on a 20 × 40 × 0.3 cm 4.8% polyacrylamide gel run at 300 V for 7 hr at 4° in 40 mM Tris-acetate (pH 7.2), 20 mM sodium acetate, and 3 mM EDTA.

Sequence Analysis. Complete digestion of microvariant RNA with either RNase T_1 or RNase A, two-dimensional electrophoretic separation of the resulting oligonucleotide mixtures, and further analysis of these oligonucleotides by digestion with an alternate nuclease or by hydrolysis with alkali were all performed according to the methods of Sanger and his colleagues (12, 13). The technique of isolating a series of partially synthesized (−) strand fragments (all possessing the 5′ end), and the method of analyzing these fragments for the presence or absence of specific oligonucleotides (fragment length mapping) has been described (10).

Partial Ribonuclease Digestion. Microvariant RNA was incubated in the presence of: 10 mM Tris·HCl (pH 7.5), 10 mM $MgCl_2$, 5 mg/ml of *E. coli* carrier RNA, and 1 μg/ml of ribonuclease T_1 at 0° for 30 min. Digestion was terminated by the addition of urea to a final concentration of 7 M. Large sequence fragments were isolated from the resulting mixture by slab gel electrophoresis on 12% polyacrylamide gels, run as described above, until a bromophenol blue dye marker moved 30 cm into the gel.

RESULTS

Microvariant RNA was first detected in MDV-1 RNA preparations that had been isolated from Qβ replicase reactions incubated for long periods of time. It gave a characteristic set of extra oligonucleotides ("ghost" spots) in autoradiographic fingerprint patterns of MDV-1 (+) RNA; and it was seen as an additional, faster-moving peak in electrophoretic analyses of MDV-1 RNA.

Microvariant RNA was isolated from a Qβ replicase reaction incubated for 120 min in the absence of an exogenous template. The RNA isolated from this reaction was a mixture of MDV-1 RNA and microvariant RNA. It was melted at 100° and separated by electrophoresis on a 4.8% polyacrylamide gel into two peaks. The larger peak contained MDV-1 RNA and the smaller, faster-moving peak contained microvariant RNA. This microvariant RNA was then used as a template for a second reaction. The resulting RNA was isolated, self-annealed at 65°, and analyzed by gel electrophoresis. The only products were microvariant (+/−) RNA and microvariant (+) RNA (which served as template for all subsequent reactions).

To determine the nucleotide sequence of microvariant RNA, four separate syntheses were performed, each containing a different ribonucleoside [α-^{32}P]triphosphate. The resulting RNAs were then digested with either RNase T_1 or RNase A, and the oligonucleotide products were separated by two-dimensional electrophoresis (12). Analyses were performed on melted (+/−) RNA, (+) RNA, and (−) RNA. Fig.

Table 1. Oligonucleotide products of the digestion of each strand of microvariant RNA with either RNase T_1 or RNase A

(+) STRAND T_1 DIGEST			(+) STRAND A DIGEST			(−) STRAND T_1 DIGEST			(−) STRAND A DIGEST		
#	MOLES	SEQUENCE	#	MOLES	SEQUENCE	#	MOLES	SEQUENCE	#	MOLES	SEQUENCE
1	2	G[G]C	51	4	Y[C]C	1	4-5	G[G]C	51	5-6	Y[C]C
	3	G[G]A		4-6	Y[C]A		2	G[G]A		3-2	Y[C]A
	2	G[G]G		2-3	Y[C]G		2	G[G]G		4-3	Y[C]G
	1	G[G]U		3	Y[C]U		2-4	G[G]U		2-1	Y[C]U
2	2	G[CG]C	52	1	Y[AC]C	2	2	G[CG]C	52	0-1	Y[AC]C
	2	G[CG]A		2	Y[AC]G		1	G[CG]G		1	Y[AC]G
3	1	G[CCG]C	53	1	Y[AAC]C		1	G[CG]U		2	Y[AC]U
	1	G[CCG]A		3	Y[GC]C	4	1	G[AG]G	53	1	Y[AAC]U
4	1	G[AG]G	55	2-3	Y[GC]A	10	1	G[UG]C	54	1	Y[AAAAC]C
	3	G[AG]U		1	Y[GC]G		1	G[UG]A		2-3	Y[GC]C
5	1	G[CAG]C	56	1	Y[GAC]G		1	G[UG]G	55	1-0	Y[GC]A
6	1	G[ACG]G	57	1	Y[AGC]G	11	1	G[UCG]C		2-3	Y[GC]G
7	1	G[CACG]U	58	2	Y[AAAGC]U	12	1	G[CUG]C		1	Y[GC]U
8	1	G[AAG]U	59	1	Y[AAAAGC]C	13	1	G[AUG]A	56	1	Y[GAC]U
9	1	G[CCAAAG]C	60	1	Y[GGC]G	14	1	G[CAUCG]C	57	1	Y[AGC]U
10	1	G[UG]C	61	1	Y[AGGGC]G	15	1	G[CCCUAG]G	59	1	Y[AAAAGC]U
	1	G[UG]G	64	1	Y[U]C	16	1	G[ACACUCCG]U	60	1	Y[GGC]A
	1	G[UG]U		4-3	Y[U]A	17	1	G[(CACUC,C)ACG]U		1	Y[GGC]G
12	1	G[CUG]A		2-1	Y[U]G	19	1	G[ACUUCG]G*		1	Y[GGC]U
13	1	G[AUG]C		7	Y[U]U	21	1	G[CUUUG]G	62	1	Y[AAAGGGC]G
18	1	G[UUG]G	65	1	Y[AU]C	23	1	G[CUUUAAAAG]C	63	1	Y[AGGAGGAC]A
20	1	G[UC(AUC,ACC)ACG]C	66	1	Y[GU]C	25	1	G[U(U,AAAACC,U,C)AG]C	64	5	Y[U]C
22	1	G[UC(C,U,C)CUAG]G*		1	Y[GU]G	26	1	G[CUUUUAAAG]G		4-2	Y[U]A
24	1	G[CCCUUUAAAAG]C	67	1	Y[GAU]G	27	1	pppG](G)		2-1	Y[U]G
26	1	G[CUUUUAAAG]C	69	1	Y[GAAGU]C	28	1	G[(CCAACUCCC)-OH		8-6	Y[U]U
27	1	pppG]G	70	2	Y[GGAGU]G				65	1	Y[AU]C
29	1	G[(UU)UUAACCC-OH	71	1	Y[GAGGU]U				66	1	Y[GU]C
			73	1	pppGGGAGU]U					2	Y[GU]G
			74	1	Y[(C)-OH				67	1	Y[GAU]G
									68	1	Y[GGU]G
									72	1	pppGG(G)U]U
									74	1	Y[(C)-OH

The sequence of each oligonucleotide is enclosed within square brackets. The letters outside these brackets indicate what was known about the nearest-neighbor nucleotides. A "Y" indicates a 5′-neighbor pyrimidine. The number of times an oligonucleotide occurs within a given strand is indicated in the "Moles" column (a range is given where this could not be determined with certainty).

Parentheses enclose uncertain portions of an oligonucleotide's sequence. Where subsequences are separated by a comma, the correct sequence may involve an interchange of the subsequences. An * after an oligonucleotide indicates that there may exist an additional 5′-terminal C.

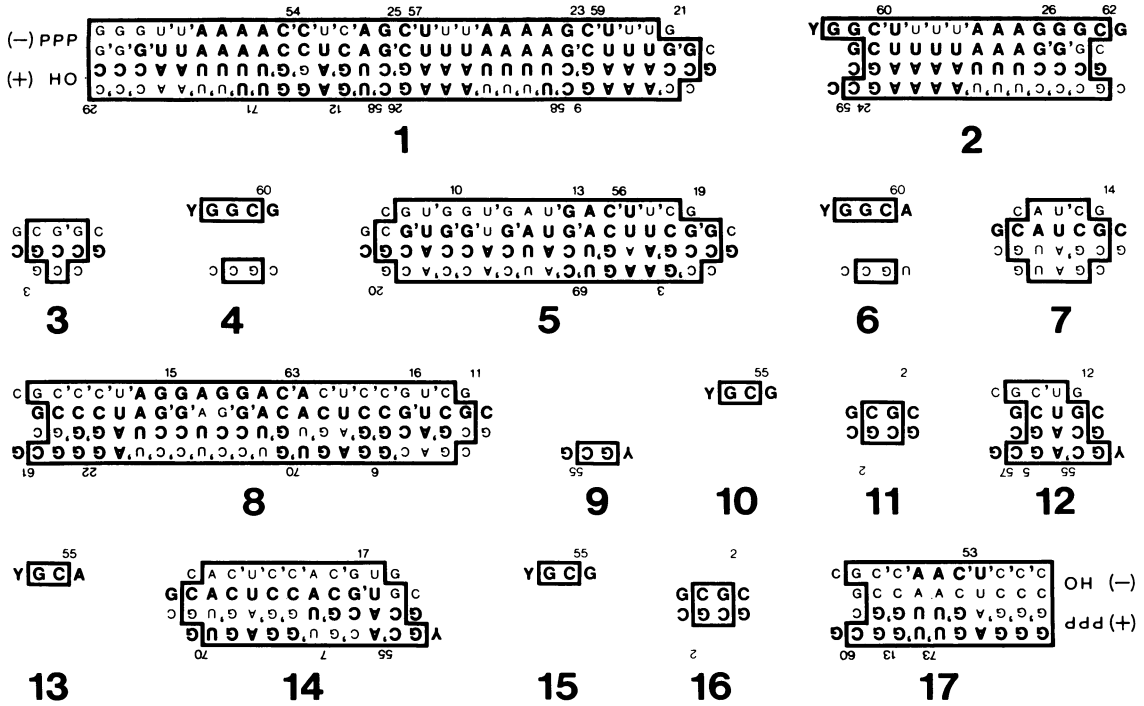

FIG. 2. Organization of the oligonucleotides identified in Table 1 into duplex blocks. A block consists of four lines, one for each of the oligonucleotide catalogs. The oligonucleotides are represented in the standard 5′ to 3′ direction. The two top lines represent the (−) strand. Since (+/−) RNA is an antiparallel duplex, the two bottom lines, representing the (+) strand, are upside down. The inner lines represent RNase T_1 oligonucleotides, and the outer lines represent RNase A oligonucleotides. The oligonucleotides assembled into each block are represented within the block outline. Apostrophes indicate the points of enzymatic cleavage. The letters outside each block outline indicate what was known about the nearest-neighbor nucleotides. Each block is identified by a large numeral. The backbone oligonucleotides are indicated by bold lettering and are identified by the small numerals outside each block outline.

1 shows the characteristic fingerprint patterns obtained from these oligonucleotide mixtures. Standard sequencing techniques (13) were used to determine the identity and frequency of occurrence of each oligonucleotide. Each RNA was labeled specifically by only one ribonucleoside, vastly simplifying the task of identifying the sequence of each oligonucleotide (10), and also providing information as to the identity of its nearest 3′-neighbor nucleotide. Table 1 shows the resulting catalog of oligonucleotides obtained for each strand of microvariant RNA digested with each of the two ribonucleases.

The sequence and multiplicity of a few oligonucleotides could not be precisely determined by these procedures. However, the availability of complementary sequence information from both the (+) and the (−) strands resolved these uncertainties. Fig. 2 shows how the information present in each of the four oligonucleotide catalogs was combined into one comprehensive description of what was known about the sequence. Each of the 17 duplex blocks represents a section of double-stranded microvariant RNA. Oligonucleotides identified in a ribonuclease digest of one strand were paired off with their oligonucleotide complements from the other strand. Similarly, overlapping oligonucleotides, identified in the two ribonuclease digests of the same strand, were also paired off. Thus, each block consists of four lines, one for each oligonucleotide catalog.

The method of block construction can be illustrated by the assembly of block 5. The sequence of oligonucleotide 20 in the RNase T_1 (+) catalog, G[UC(AUC,ACC)ACG]C, was uncertain. Its exact sequence was established by pairing it off with two smaller oligonucleotides found in the complementary RNase T_1 (−) catalog. The unique RNase T_1 (−) oligonucleotides 10, G[UG]G, and 13, G[AUG]A, could only form a complement with RNase T_1 (+) 20; and, furthermore, they could form this complement with only one of the two alternative sequences of RNase T_1 (+) 20. The complement of the 5′ end of RNase T_1 (+) 20 is UGAC. There was only one U[GAC] in the RNase A (−) catalog, namely, fragment 56, Y[GAC]U, and it was added to the block. Its complement is AGUC. Searching the RNase A (+) catalog for oligonucleotides terminating AGU]C, only one was found, namely fragment 69, Y[GAAGU]C, and it was also added to the block. In a similar manner, RNase T_1 (−) 19 was added, followed by RNase T_1 (+) 3. At this point no further extension of the block was possible. For example, an RNase A (+) oligonucleotide terminating GC]C would extend the block; however, there were two different oligonucleotides of this

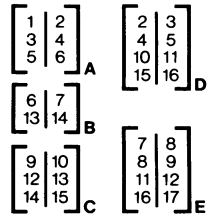

FIG. 3. Restrictions governing the joining together of the duplex blocks. Each of these five logical choice sets identifies two groups of blocks: members of the first group, listed to the left of the vertical line, possess similar right sides; and members of the second group, listed to the right of the line, possess similar left sides. Each member of the group on the left can form a perfect match with each member of the group on the right. For example, choice set E indicates that the right sides of blocks 7, 8, 11, and 16 match the left sides of blocks 8, 9, 12, and 17.

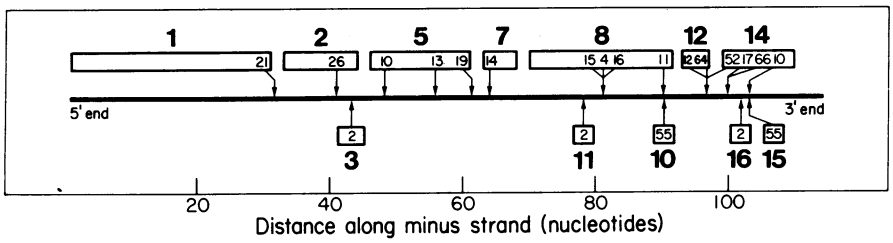

FIG. 4. Fragment length map of guanosine-labeled microvariant (−) RNA. The heavy line in the center of the diagram represents the minus strand. Arrows indicate the location at which each oligonucleotide was mapped. Oligonucleotides are identified by the small numerals, and the blocks that contain them are indicated by rectangles and identified by the large numerals.

type, namely, RNase A (+) 55 and 59, and a choice could not be made between them. Once the "backbone" oligonucleotides were assembled, other oligonucleotides complementary to the backbone elements were added to the block. Although these smaller oligonucleotides were not essential to block assembly, their exact multiplicities were determined, and in each case they fell within the experimental range. For example, the experimental data indicated that two or three Y[GC]C fragments would occur within the (−) strand. However, only two of these oligonucleotides were required during block construction, and they were placed in blocks 8 and 17.

The duplex blocks could be joined together into a complete representation of (+/−) RNA. Significantly, there were restrictions as to which blocks could be linked with a given block. Two blocks could be joined together only if the letters outside the outline of one block (representing nearest-neighbor nucleotides) matched the letters within the outline of the other block (representing terminal nucleotides). Fig. 3 summarizes these restrictions in the form of "logical choice sets." They substantially limit the number of possible sequence permutations. The task of determining the final nucleotide sequence was thus reduced to finding the nearest linkage relationships within each choice set.

These relationships were determined experimentally by the construction of a fragment length map of microvariant RNA (10). This technique involves the synthesis and isolation by gel electrophoresis of a series of increasingly longer microvariant (−) strand fragments, each containing the 5′ end. Eighteen size classes varying in length from 29 to 114 nucleotides were isolated and examined. Each fragment was fingerprinted in order to determine the minimum length of RNA required before a given oligonucleotide could be detected. Key oligonucleotides, and consequently the blocks in which they occur, were mapped in this manner. Fig. 4 summarizes the data obtained from these experiments. Oligonucleotides contained in the same block were found close together on the map. The following order was found for the larger blocks: 1, 2, 3, 5, 7, 8, 12, 14. In addition, the smaller blocks, 10, 11, 15, and 16, all occurred after block 7 on the map. These two observations were sufficient to "solve" the choice sets. For example, in choice set B (Fig. 3), block 14 mapped after block 7. From choice set A it was known that block 6 joins with either block 1, 3, or 5, and these three blocks mapped prior to block 7. Therefore, in choice set B, block 6 must join with block 7, and block 13 (which could not be mapped with guanosine-labeled RNA) must join with block 14. Thus, the mapping data, in conjunction with the logical restrictions imposed by the choice sets, were sufficient to permit the conclusion that the nucleotide sequence consisted of blocks 1 through 17 in numerical order. The only ambiguity was that the small contiguous subsequence formed by blocks 9-10-11 could occur immediately before or immediately after block 8. Fortunately, an oligonucleotide fragment, isolated by slab gel electrophoresis from a partial RNase T_1 digest of (+) RNA, was found to contain blocks 9 through 17, but not block 8. Thus, the 9-10-11 subsequence occurs after block 8. Fig. 5 illustrates the complete nucleotide sequence of microvariant (+) RNA.

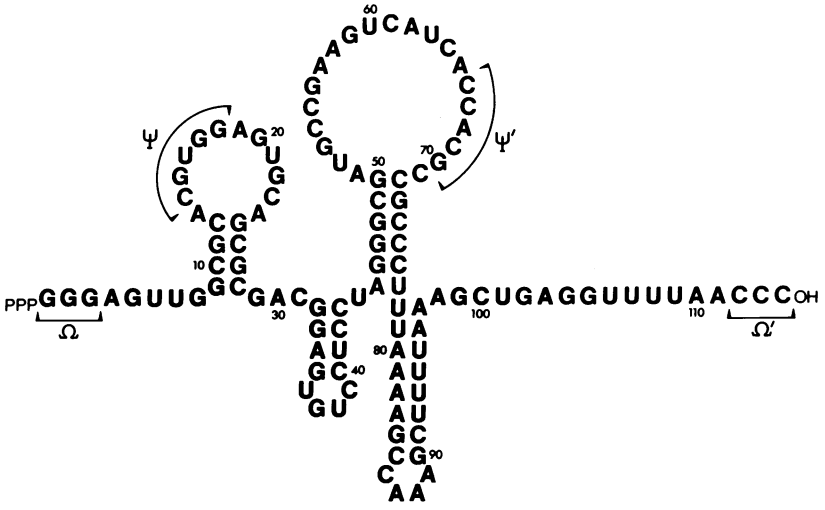

FIG. 5. The complete nucleotide sequence of microvariant (+) RNA. Nothing is known yet about the secondary structure of this molecule. However, many regions of intrastrand complementarity are present. Four of the potentially strongest complements (16) are illustrated in the form of "hairpin" secondary structures. Two other potentially strong intrastrand complements occur in the regions identified as Ψ and Ψ′ and Ω and Ω′.

DISCUSSION

Several different small RNA molecules have been isolated from Qβ replicase reactions incubated in the absence of exogenous template (17, 18), including MV-1 RNA (19), MDV-1 RNA (9), microvariant RNA, and an RNA of 90 nucleotides in length that has also recently been isolated and sequenced (W. Schaffner and C. Weissmann, personal communication). All are recognized by Qβ replicase and are synthesized in a manner similar to Qβ RNA. Since the complete nucleotide sequences of at least two are available, it becomes feasible to compare them for possible common features.

Both microvariant RNA and MDV-1 RNA have a short oligo(G) stretch at their 5′ ends, and a complementary oligo(C) stretch at their 3′ ends (a feature found in many viral RNAs not copied by Qβ replicase). Also, both RNAs have fewer AU dinucleotides than would be predicted on the basis of their nucleotide compositions. Furthermore, both RNAs contain many different intrastrand complements, suggesting the existence in both of a highly structured single-strand conformation. However, they have a dissimilar G-C content, and an examination of their respective oligonucleotide catalogs does not reveal any similarities among the larger oligonucleotides. It is clear, therefore, that a unique primary sequence is not required as a key element in the recognition and replicative processes. The implication is that secondary or tertiary structures play such roles. Future experiments can now be based on a comparison of two different self-replicating molecules, both of known nucleotide sequence.

We thank Dr. Rivka Rudner for her critical reading of the manuscript. We also thank John Mack and Edward Hajjar for the preparation of the illustrations. This work was supported by National Science Foundation Research Grant GB-17251X2 and National Institutes of Health Research Grant CA-02332.

1. Haruna, I. & Spiegelman, S. (1965) *Proc. Nat. Acad. Sci. USA* **54,** 579–587.
2. Haruna, I. & Spiegelman, S. (1965) *Science* **150,** 884–886.
3. Spiegelman, S., Haruna, I., Holland, I. B., Beaudreau, G. & Mills, D. (1965) *Proc. Nat. Acad. Sci. USA* **54,** 919–927.
4. Pace, N. R. & Spiegelman, S. (1966) *Science* **153,** 64–67.
5. Mills, D. R., Peterson, R. L. & Spiegelman, S. (1967) *Proc. Nat. Acad. Sci. USA* **58,** 217–224.
6. Levisohn, R. & Spiegelman, S. (1968) *Proc. Nat. Acad. Sci. USA* **60,** 866–872.
7. Levisohn, R. & Spiegelman, S. (1969) *Proc. Nat. Acad. Sci. USA* **63,** 805–811.
8. Saffhill, R., Schneider-Bernloehr, H., Orgel, L. E. & Spiegelman, S. (1970) *J. Mol. Biol.* **51,** 531–539.
9. Kacian, D. L., Mills, D. R., Kramer, F. R. & Spiegelman, S. (1972) *Proc. Nat. Acad. Sci. USA* **69,** 3038–3042.
10. Mills, D. R., Kramer, F. R. & Spiegelman, S. (1973) *Science* **180,** 916–927.
11. Kramer, F. R., Mills, D. R., Cole, P. E., Nishihara, T. & Spiegelman, S. (1974) *J. Mol. Biol.* **89,** 719–736.
12. Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13,** 373–398.
13. Brownlee, G. G. (1972) in *Determination of Sequences in RNA* (American Elsevier Publishing Co., New York), pp. 1–265.
14. Eoyang, L. & August, J. T. (1971) in *Procedures in Nucleic Acid Research,* eds. Cantoni, G. L. & Davies, D. R. (Harper & Row, New York), Vol. 2, pp. 829–839.
15. DeWachter, R. & Fiers, W. (1971) in *Methods in Enzymology,* eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 21, part D, pp. 167–178.
16. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature New Biol.* **246,** 40–41.
17. Banerjee, A. K., Rensing, U. & August, J. T. (1969) *J. Mol. Biol.* **45,** 181–193.
18. Sumper, M. & Luce, R. (1975) *Proc. Nat. Acad. Sci. USA* **72,** 162–166.
19. Kacian, D. L., Mills, D. R. & Spiegelman, S. (1971) *Biochim. Biophys. Acta* **238,** 212–223.