



Universiteit  
Leiden  
The Netherlands

## Random removal of inserts from an RNA genome: selection against single-stranded RNA

Olsthoorn, R.R.C.L.; Duin, J. van

### Citation

Olsthoorn, R. R. C. L., & Duin, J. van. (1996). Random removal of inserts from an RNA genome: selection against single-stranded RNA. *Journal Of Virology*, 70(2), 729-736.  
doi:10.1128/jvi.70.2.729-736.1996

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/3631969>

**Note:** To cite this publication please use the final published version (if applicable).

## Random Removal of Inserts from an RNA Genome: Selection against Single-Stranded RNA

RENÉ C. L. OLSTHOORN AND JAN VAN DUIN\*

*Gorlaeus Laboratories, Department of Biochemistry, Leiden Institute of Chemistry,  
University of Leiden, 2300 RA Leiden, The Netherlands*

Received 15 August 1995/Accepted 24 October 1995

**We have monitored the evolution of insertions in two MS2 RNA regions of known secondary structure where coding pressure is negligible or absent. Base changes and shortening of the inserts proceed until the excessive nucleotides can be accommodated in the original structure. The stems of hairpins can be dramatically extended but the loops cannot, revealing natural selection against single-stranded RNA. The 3' end of the MS2 A-protein gene forms a small hairpin with an *Xba*I sequence in the loop. This site was used to insert *Xba*I fragments of various sizes. Phages produced by these MS2 cDNA clones were not wild type, nor had they retained the full insert. Instead, every revertant phage had trimmed the insert in a different way to leave a four- to seven-membered loop to the now extended stem. Similar results were obtained with inserts in the 5' untranslated region. The great number of different revertants obtained from a single starting mutant as well as sequence inspection of the crossover points suggest that the removal of redundant RNA occurs randomly. The only common feature among all revertants appears the potential to form a hairpin with a short loop, suggesting that single-stranded RNA negatively affects the viability of the phage. To test this hypothesis, we introduced *Xba*I fragments of 34 nucleotides that could form either a long stem with a small loop or a short stem with a large loop (26 nucleotides). The base-paired inserts were perfectly maintained for many generations, whereas the unpaired versions were quickly trimmed back to reduce the size of the loop. These data confirm that single-stranded RNA adversely affects phage fitness and is strongly selected against. The repair of the RNA genome that we describe here appears as the result of random recombination. Of the plethora of recombinants, only those able to adopt a base-paired structure survive. The frequency with which our inserts are removed seems higher than measured by others for small inserts in a reading frame in Q $\beta$  RNA. To account for this higher frequency, we suggest models in which the single-stranded nature of our inserts induces random recombination at the site of the insertion.**

The single-stranded RNA coliphages are among the smallest known viruses. Their compact genomes vary between 3,500 and 4,200 nucleotides (nt) in size and code for four proteins which take up about 90% of the space. The other 10% is noncoding but contains important elements for replication and translation. The RNA is highly base paired and adopts a specific secondary structure, much of which is conserved among the different phage groups (2, 23, 29, 30). Secondary structure is assumed to play an important role in many aspects of the life cycle of the phage. Apart from protecting against degradation by RNases and condensing the RNA to fit the virus shell, secondary structure also regulates translational initiation, provides binding sites for viral proteins, and facilitates strand separation during replication (1, 27, 31).

Because of the lack of proofreading by the viral replicase, reproduction has a high error rate, and misincorporation frequencies of  $10^{-3}$  to  $10^{-4}$  are typical (7). As a consequence, the genome is a mixture of related sequences which is referred to as a quasi-species (5, 9). Natural selection is likely to maintain the consensus sequence, which explains why, for example, Q $\beta$  isolates have hardly changed during several decades of laboratory culture (14, 30).

RNA viruses are able to repair distortions in their genomes by a variety of mechanisms which include reversion, pseudoreversion, and recombination. In the past, the high reversion

frequency in RNA phages has masked the occurrence of recombination, and for a long time it was thought that RNA recombination did not occur in these creatures (12). However, the discovery of autocatalytic templates of Q $\beta$  replicase composed of heterologous sequences was a strong indication for the existence of recombination (18–20). More recently, in vitro RNA recombination by Q $\beta$  replicase was demonstrated when a self-replicating RNA (SV-11) could be generated from another (MNV-11) by a duplication event (3). Homologous recombination was proposed to account for the reversion of Q $\beta$  deletion and insertion mutants in the replicase gene (26). In these experiments, the repair of the Q $\beta$  replicase gene always resulted in wild-type recombinants because of the strong selection pressure on the coding sequence. Distortions of reading frames are therefore not generally suitable to reveal other recombinants or other types of selection pressure, for instance, those acting on the secondary structure of the RNA.

To examine this point in more detail, we have made insertions in regions of the MS2 genome where selection pressure on the reading frame is weak or absent and where we know the secondary structure from probing data and phylogenetic comparison (10, 29). Using an infectious DNA copy of MS2, we introduced fragments ranging in size from four to several hundreds of nucleotides in the 3'-terminal nucleotides of the A-protein gene and in the 5' noncoding region. These inserts decrease the titer of the infectious clone by several orders of magnitude. Passaging of the recombinants raises the titer in parallel with the reduction and molding of the insert. Inspection of the remnant sequences in revertants shows that, in general, excision is not guided by any sequence homology but

\* Corresponding author. Mailing address: Leiden Institute of Chemistry, Department of Biochemistry, Gorlaeus Laboratories, University of Leiden, P.O. Box 9502, 2300 RA Leiden, The Netherlands. Phone: (31) 71-5274759. Fax: (31) 71-5274340.

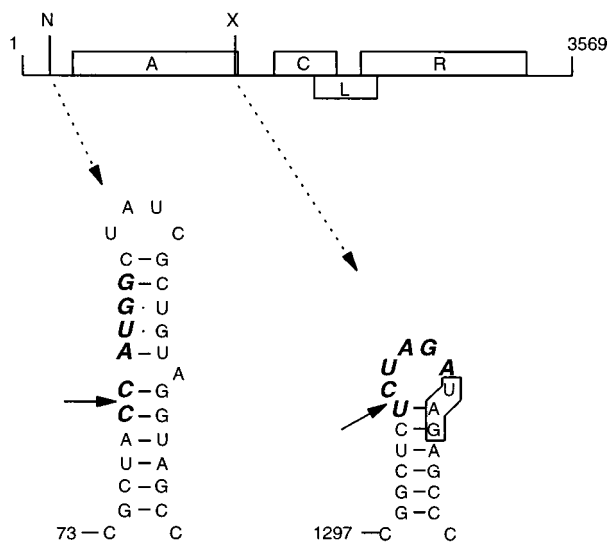


FIG. 1. Genetic map of bacteriophage MS2 showing sites used for inserting extra nucleotides. A, C, L, and R represent the genes for the A (or maturation), coat, lysis, and replicase proteins, respectively. N and X indicate restriction sites for *Nco*I and *Xba*I, present in MS2 cDNA. The RNA secondary structure at these positions was deduced by Groeneveld et al. (10) and Skripkin et al. (28). Sequences corresponding to the *Nco*I and *Xba*I sites are presented in boldface italics. The stop codon of the A gene is boxed.

appears as a random event. Furthermore, inserts are not removed when the RNA is predicted to adopt a base-paired structure. The experiments show that selection pressure acts against long regions of single-stranded RNA.

## MATERIALS AND METHODS

**Bacterial strains.** All plasmids in this study carry the  $p_L$  promoter of phage  $\lambda$  and were grown in *Escherichia coli* K-12 strain M5219 [M72 lacZ(Am) *trp*(Am) Sm<sup>r</sup> ( $\Delta$ bio252c1857  $\Delta$ H1)], encoding the thermosensitive  $\lambda$  repressor (cI857) and the transcription antitermination factor N (28). Evolution of phages was allowed to occur in *Escherichia coli* F<sup>+</sup> strain GM-1 (F' *lacI<sup>s</sup>*, *pro/ara*  $\Delta$ lac-*pro* *thi*) (16). Both strains were grown on LC broth (24).

**Insertions in the *Xba*I site at position 1303.** Insertions were made in the infectious cDNA clone pMS2000, the construction of which will be described elsewhere, or in plasmids pMS2029 and pMS2031, which are derivatives of pMS2000 that carry the silent marker mutations U-1361→C and A-1379→G; pMS2029 has an additional U-1373→C substitution. These substitutions are stably maintained in the MS2 phage (24). The *Xba*I site, position 1303 in the MS2 genome, is located directly upstream of the stop codon of the A-protein gene (Fig. 1). Mutant plasmids pMS.L and pMS.M were obtained by inserting *Xba*I fragments of ~300 and 122 bp in pMS2029 and pMS2031, respectively. These fragments were picked up during the construction of pMS2000 mutants and probably derive from the *E. coli* chromosome. The complete sequence of the 122-bp fragment was determined, and the 300-bp insert was partially sequenced. No sequence homology with any entry in the GenBank (release 87.0) was found for either fragment. The sequence of the pMS.M insert (a potential stop codon is underlined) is

```
CTAGAACCATTATCTAAAGTTGAGCCAAAAGAGCTCAAGTTGAGTATTTTCTAAAAGTC
ACGTGGAAATAATAAGATTAGACAGTTTGTATGGAAATAAATGCAGATTTTAAAATAAC
TT
```

The sequence of the pMS.L insert is

```
CTAGAGTTATTTCTTCTCTCCTCATATTCACAAATTATCTTATGATGCTATCATAATTGAAA
GATTAGCCAGGGACCCATATCATGGCTATGGCAGTCATGTCCTGTGAAAAATCCATTAAA
GGACCTCC...~100 bp...GGACTTTCTCCTGTGTACCTTCTCAGTGGAGAGTCCA
CCTGCTGGTAGTGCACCTCCTTAGGTGCTGTCTGGGCTGTCTGTGCCACCTGCACAG
AGTCATGCATGAGCTCCACTGTGGGAAGTCT
```

For construction of the 34- and 68-bp insertion mutants, the following pairs of oligonucleotides were used: A (5'CTAGATAGAAGCTCCTAACTAAACAACCTA ACCAAAT) and B (5'CTAGATTGGTTAGTTGTTAGTTAGGAGTTCT

AT); and C (5'CTAGATAGAAGCTGCTAGCTAACTAGCTGTTCTTT) and D (5'CTAGAAAAGAAGCTAGTTAGCTAGCTAGTTCTAT) (bases in boldface mark an *Nhe*I restriction site). Phosphorylated oligonucleotides C and D were heated for 5 min at 90°C in 20  $\mu$ l of ligation buffer and slowly cooled to room temperature. Subsequently, *Xba*I-digested pMS2031, ATP, and T4 DNA ligase (New England Biolabs) were added, and the total mixture incubated overnight at 14°C. The ligation mixtures were precipitated with ethanol, and the pellets were dissolved in distilled water. About one-quarter of this solution was transformed to M5219 cells by means of electroporation (12.5 kV/cm; Cellject). Kanamycin-resistant transformants were tested for phage production by replicating on a lawn of GM-1 F<sup>+</sup> cells. About two-thirds of the transformants produced phages. From 24 of these, DNA was isolated and incubated with *Nhe*I. Clones containing an *Nhe*I site were analyzed by DNA sequencing. Both orientations of the monomeric inserts, pMS.C (with C in the coding strand) and pMS.D (with D in the coding strand), as well as the double inserts pMS.E and pMS.G, which contain duplications of C and D, respectively, were obtained.

Plasmids pMS.A and pMS.B (with oligonucleotides A and B, respectively, in the coding strand) were obtained in the same way except that *Xba*I-digested pMS2031 was simply coprecipitated with nonphosphorylated oligonucleotides A and B before ligation.

Phages generated by cells harboring pMS.A, pMS.B, pMS.C, pMS.D, pMS.E, or pMS.G are designated A.n, B.n, C.n, D.n, E.n, or G.n, respectively, where n = 0 specifies the starting mutants. Their evolutionary offspring are indicated as A.1, A.2, etc. This numbering is arbitrary and does not necessarily imply an evolutionary route. Likewise, phages obtained from pMS.L and pMS.M are identified by the letters L and M, respectively. Sometimes, the single-letter code is used for simplicity to indicate the starting mutant.

**Insertions in the 5' UTR.** Insertion mutants in the 5' untranslated region (UTR) were made via pPLAMS22, which was originally constructed for expression studies on the maturation gene (10). This plasmid contains the MS2 sequence from positions 1 to 1303 and confers kanamycin resistance. A 99-bp *Afl*III fragment from gene I of M13 (positions 3619 to 3717) was inserted in both orientations in the *Nco*I site of pPLAMS22 (position 78 of the MS2 genome). The two resulting clones are designated pHG227 and pHG229; the latter contains the gene I fragment in the sense orientation. Phages with inserts were obtained by transforming pHG227- or pHG229-containing cells with pPLC236/2, which carries MS2 cDNA information from position 103 to the 3' end and the *bla* gene for ampicillin resistance. Because of the considerable overlap in MS2 information between the two plasmids (the region from position 103 to 1303), genetic recombination at the DNA, but more likely at the RNA, level can occur, giving rise to the desired mutant phages. This procedure will be described in more detail elsewhere. Transformants resistant to both ampicillin and kanamycin were grown in 2.5 ml of LC plus both antibiotics. After overnight growth at 28°C, 10  $\mu$ l of supernatant from each culture was screened for the presence of phages. In these aliquots, no phage could be detected. The overnight cultures were therefore pooled and treated with lysozyme. From pHG227/pPLC236 cells, approximately 1 PFU/ml of culture was obtained. pHG229/pPLC236 cells produced <0.5 PFU/ml. Control experiments with the wild-type pPLAMS22 vector (positions 1 to 1303) typically yielded titers of 10<sup>6</sup> PFU/ml, indicating that recombination does indeed occur and that the M13 insert in either orientation drastically reduces phage production. Phages derived from pHG227/pPLC236 and pHG229/pPLC236 cells are designated P.1, P.2, etc., and Q.1, Q.2, etc., respectively.

**Sequence analysis.** Plasmid DNA was sequenced by the dideoxynucleotide chain termination method, using a T7 sequencing kit (Pharmacia). Phage RNA was isolated and sequenced by reverse transcription as described before (24). For the 5' UTR mutants, an oligonucleotide (DUI360) complementary to bases 146 to 167 was used. Mutations in the *Xba*I site were sequenced by using an oligonucleotide complementary to bases 1409 to 1422 (DUI59).

**Recording phage evolution.** The procedure to monitor phage evolution has been described in detail elsewhere (24, 25). Briefly, mutant phages were produced by overnight growth at 28°C of *E. coli* M5219 containing a plasmid with mutant MS2 cDNA. Part of the supernatant was plated on lawns of GM-1 F<sup>+</sup> cells. Phages derived from single plaques were amplified in a fresh liquid culture of GM-1 cells to obtain sufficient material for sequence analysis. We define growth in M5219 cells as cycle 1, the first growth as plaques in GM-1 cells as cycle 2, and the first growth in liquid cultures of GM-1 cells as cycle 3. Subsequent infection cycles started with the inoculation of GM-1 cultures with 1  $\mu$ l (10<sup>6</sup> to 10<sup>8</sup> phage) of supernatant of the previous cycle and ended after overnight growth at 37°C. The multiplication factor of one cycle is about 1,000.

## RESULTS

**The system.** A complete cDNA copy of phage MS2 or a mutant thereof is present on a plasmid under transcriptional control of the thermoinducible promoter  $p_L$  from phage  $\lambda$ . The plasmid is maintained in a bacterial strain that does not produce F pili and thus cannot be reinfected from without. Overnight growth of a bacterial culture containing the plasmid yields phage (even without induction of the promoter) which are released into the medium. This supernatant is diluted and

TABLE 1. Properties of wild-type and mutant cDNA clones

MS2 cDNA plasmid (phage)	Titer (PFU/ml)	Length of insert (bp)
pMS2000 (wild type)	$1 \times 10^{11}$	
pMS.R (R) <sup>a</sup>	$5 \times 10^{10}$	4
pMS.L (L) <sup>b</sup>	$2 \times 10^7$	~300
pMS.M (M) <sup>c</sup>	$1 \times 10^7$	122
pMS.A (A) <sup>c</sup>	$2 \times 10^8$	34
pMS.B (B) <sup>c</sup>	$2 \times 10^8$	34
pMS.C (C) <sup>c</sup>	$2 \times 10^{10}$	34
pMS.D (D) <sup>c</sup>	$3 \times 10^{10}$	34
pMS.E (E) <sup>c</sup>	$3 \times 10^9$	68
pMS.G (G) <sup>c</sup>	$7 \times 10^9$	68
pHG227/pPLc236 (P)	$1 \times 10^{5d}$	99
pHG229/pPLc236 (Q)	$5 \times 10^{4d}$	99

<sup>a</sup> Constructed by fill-in of the *Xba*I site of pMS2000.

<sup>b</sup> Has U-1361→C and A-1379→G as silent markers and carries the additional substitution U-1373→C.

<sup>c</sup> Has U-1361→C and A-1379→G as silent markers.

<sup>d</sup> Titers were normalized to a one-plasmid system (see Materials and Methods).

plated on lawns of GM-1 cells, which carry F pili. Evolution is then further pursued in liquid medium by inoculation of GM-1 with phages derived from a single plaque as described in Materials and Methods section. The wild-type MS2 cDNA plasmid produces about  $10^{11}$  PFU per ml of culture during overnight growth.

**Insertions in the 3' end of the A-protein gene.** To test the tolerance and flexibility of the MS2 RNA genome, we made insertions in the 3' end of the A-protein gene (Fig. 1). This region forms a small hairpin structure with an *Xba*I sequence in the loop (29). Via the infectious MS2 cDNA clone, *Xba*I fragments of ~300 and 122 bp were inserted. The 122-nt insert is AU rich and was predicted by MFOLD (13) not to rearrange the MS2 RNA folding but to be present as a structure of its own on top of the terminator hairpin. The resulting plasmids, pMS.L and pMS.M, encode phage mutants L and M, respectively. The inserts caused a  $10^3$ - to  $10^4$ -fold reduction in the titer of the infectious clone (Table 1).

Several randomly picked plaques obtained from pMS.L and pMS.M were amplified in liquid cultures of F<sup>+</sup> cells. After overnight growth at 37°C, these lysates had again titers of about  $5 \times 10^{11}$  PFU/ml, similar to that of wild-type phages. Phage RNA from 11 descendants was characterized, and their sequences at the site of the insertion are shown in Fig. 2. None of the phages had retained the insert, nor had any developed the wild-type sequence, although this can be achieved by homologous recombination at the *Xba*I site on a repeat of six and eight bases in mutants L and M, respectively. Instead, in the individual phages, the insert had been reduced to different extents. There is no apparent relationship between size or sequence of the insert and the end products, but all of the inserts are trimmed back to 13 or fewer nt. From the viewpoint of RNA secondary structure, the size and choice of the deleted sequence can, however, be readily understood. All deletions allow the revertants to form a hairpin with a loop of 4 to 7 nt (Fig. 3). This size corresponds to what is statistically preferred in hairpin loops in phage RNA (our unpublished observation). In most of the revertants, the *Xba*I repeat serves to extend the stem. The presence of extra nucleotides stored in the extended stem-loop structure apparently does not affect phage viability as judged by the titer. An independent confirmation for this observation is that a 4-nt duplication in the *Xba*I site (pMS.R) hardly affected infectivity of the cDNA clone (Table 1). The resulting phage maintained the 4-nt insert.

**Insertions in the 5' UTR.** A similar set of experiments was carried out with insertions in the 5' UTR. This region folds into a cloverleaf structure, which by the kinetics of its folding regulates translation of the A-protein gene (10). A 99-bp *Afl*III fragment derived from M13 was cloned in either orientation in the *Nco*I site at nt 78 of MS2 cDNA in the 5' UTR. This position is found in the side of one of the three stem-loop structures, which together build the cloverleaf (Fig. 1). Both inserts lead to overproduction of the A-protein by disrupting the formation of the cloverleaf (reference 10 and unpublished results).

Mutant phages were generated as described in Materials and Methods. From a total of 12 plaques that were sequenced, we identified five types of revertants (Fig. 4). As before, most of the inserted sequence disappeared. Revertants P.1, P.2, and Q.2 were stable for six cycles, but two cycles later, the wild type had taken over in these phage lines (other revertants were not tested). As found above, none of the cDNA clones yielded wild type initially.

We examined whether the P and Q revertants could still form a hairpin structure. The extra nucleotides in the Q series can be accommodated as small bulges (not shown). The inserts retained in the P series are bigger. Our proposal for the secondary structure assumed by P.2 is shown in Fig. 5. The extra A→C suppressor mutation that we have observed in this revertant makes sense since it allows stem formation. Without this substitution, a loop much larger than 10 nt would be predicted. The deletion in P.1 can be rationalized in a similar way (not shown).

**Single-stranded RNA reduces viral fitness.** It remains intriguing that among a total of 23 revertants analyzed, no wild type developed in the initial stages of evolution, although it can simply be generated by homologous recombination at the short repeat of the restriction site. It would seem, therefore, that the removal of extra sequences occurs quite randomly. Countless numbers of recombinants are probably formed, but selection for the formation of a hairpin with a short loop reveals only those that are viable or win the competition with less fit spec-

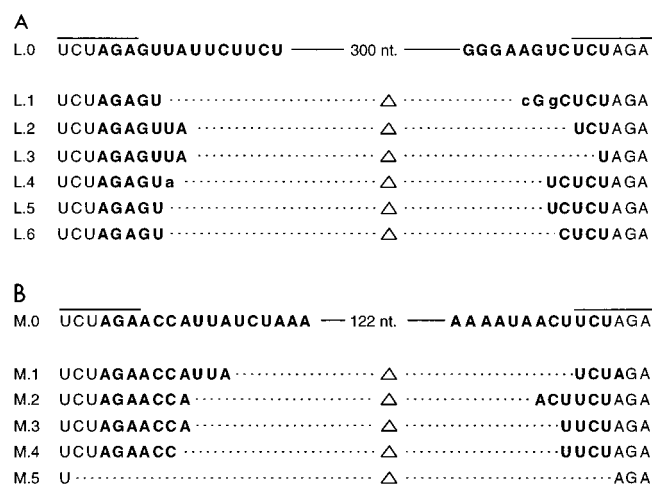


FIG. 2. Revertants obtained from mutants containing insertions in the 3' end of the A-protein gene. Phages were analyzed after three cycles of evolution (see Materials and Methods for definition of cycle number). The insert is shown in boldface. Only nucleotides directly bordering the *Xba*I sequences (overlined) are shown. Δ represents the deleted bases. (A) Reversion mutants from L, originally containing an insert of approximately 300 nt at position 1303 (*Xba*I site). Mutated nucleotides in L.1 and L.4 are indicated in lowercase. (B) Reversion mutants from M, originally containing an insert of 122 nt.

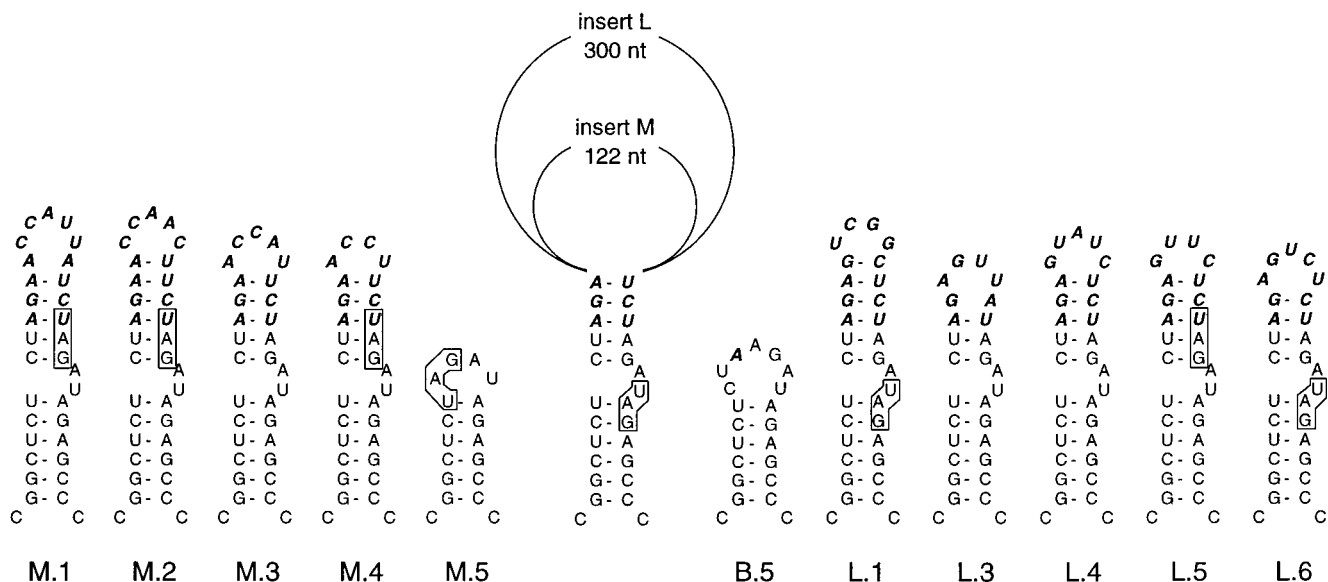


FIG. 3. Stem-loop structures predicted to form by revertants L.1 to L.6 and M.1 to M.5. L.2 is identical to L.5 but has a GUUA loop sequence. The remnants of the inserted sequence are shown in boldface italics. The A-protein stop codon in each revertant is boxed. In revertants M.3, L.3, L.4, and B.5, the stop codon lies 11 nt into the coat protein gene. B.5 is one of the revertants from mutant B.

imens. All viable recombinants have the ability to properly accommodate the excess RNA in a secondary structure. We therefore suppose that there is a strong selection against single strandedness.

To test our ideas about the harmful effects of single strandedness, we constructed two types of insertion mutants in the *Xba*I site: those that can form a long stem with a small loop (mutants C and D; Fig. 6) and those that are predicted to form a short stem with a large loop (mutant A [Fig. 7] and mutant B [not shown]). Mutant B adopts the same structure as mutant A but carries the complementary loop sequence. We have introduced two mismatches in the extended hairpins of mutants C and D because long regular hairpins do not occur in phage RNA, probably because such structures are a substrate for RNase III (4, 6, 11). All four mutants carry silent markers (see Table 1, footnote c) which allow us to distinguish between true revertants and contamination by wild-type MS2 during further growth.

The effect of the inserts on the infectivity of the cDNA clone appears correlated with the RNA structure that is predicted to form. Mutants C and D, which can form a stem, only show a three- to fivefold reduction in infectivity compared with the wild type, while for mutants A and B, phage production is

P.0	CCAUGUAAUUU	99 nt.	UUCUGUCGAGCAGCAGCAUAUAAACAACAUUGG
P.1 (3)	CCAUGUA	△	GACGcCAAUAAACAACAUUGG
P.2 (4)	CC	△	GUCCAGCAGCAGcCAAUAAACAACAUUGG
Q.0	CCAUGUGUUU	99 nt.	CUGCCUAAUUACAUGG
Q.1 (3)	CCAUGU	△	UGG
Q.2 (1)	CCA	△	UACAUGG
Q.1 (1)	CCA	△	AUACAUGG

FIG. 4. Revertants from insertions in the 5' UTR after four cycles of evolution. The number of plaques sequenced for each revertant is indicated in parentheses. P.0 and Q.0 are the input mutants. Both revertants from P.0 have sustained an A→C mutation, shown in lowercase.

decreased by 3 orders of magnitude (Table 1). This finding clearly indicates that in contrast to single-stranded RNA, the base-paired versions of the inserts do not affect any vital functions of the phage. A possible adverse effect on translation of the downstream coat protein gene by the single-stranded inserts is unlikely. In all four cDNA constructs (A to D), coat protein production was not significantly different from that of the wild type (pMS2000) (data not shown).

**Evolution of the unpaired mutants A and B.** About half of the plaques derived from pMS.A had a normal size (~1 mm), whereas the other half were unusually small (≤0.2 mm). To

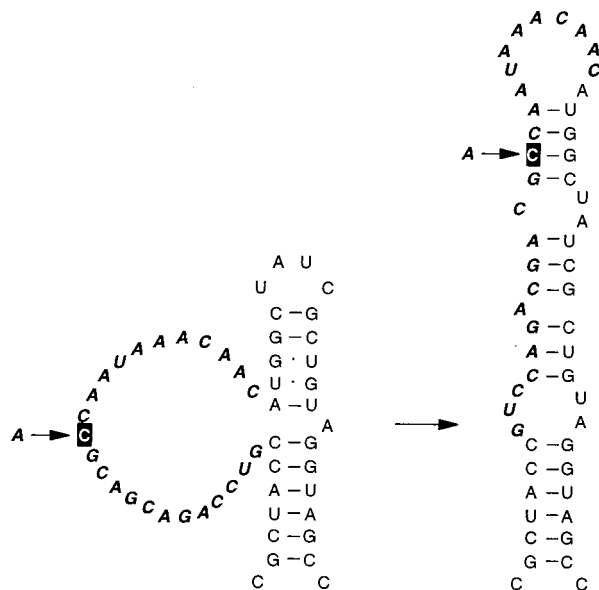


FIG. 5. Proposed folding of revertant P.2 (right). Nucleotides remaining from the initial insert (shown in boldface italics) can be accommodated in the proposed hairpin. The indicated A→C substitution supports the new folding.

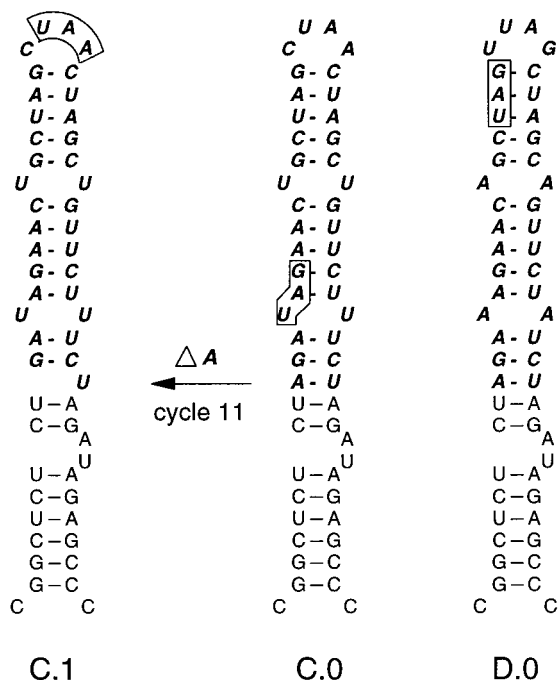


FIG. 6. Insertion mutants C.0 and D.0 which can form stable stem-loop structures. The A protein encoded in mutant C.0 is the same as the wild type; the one encoded in mutant D.0 is three amino acids longer. See the legend to Fig. 3 for further details.

test whether there was any correlation between plaque size and genotype, two normal-size and two small plaques were grown on GM-1 cells in liquid medium, and phage RNA was isolated and sequenced. The two normal-size plaques contained the wild-type MS2 sequence at the *Xba*I site. Since they carried the marker mutations, they were true recombinants.

Phages obtained from the two small plaques (A.1 and A.2) still contained the insert but had acquired a few substitutions in the loop, most of which were difficult to interpret, but we assume that they somehow increased loop structure. One of these phages (A.1; Fig. 7) quickly reverted to wild type (plus markers) upon further growth. The other, revertant A.2, was slowly overgrown by A.3, which is a typical revertant with a 5-nt loop. For reasons we do not understand, it had an additional C→U mutation in the stem (Fig. 7).

This experiment demonstrates that there is a correlation between plaque size and genotype. The large loop is related to the tiny-plaque phenotype, and subsequent removal of this large loop coincides with an increase in plaque size. The large loop somehow impairs phage reproduction and is thus strongly selected against.

In contrast to the previous revertants P, Q, L, and M, half of the revertants of mutant A were pseudo-wild type. This may be related to the fact that the repeat in mutant A is 10 nt, whereas it is at most 8 nt in the previous mutants, e.g., mutant L. This difference may explain the increased incidence of homologous recombination in mutant A leading to the wild-type sequence.

Five random plaques generated by pMS.B were selected for analysis. The repeat at the *Xba*I site in this series is 7 nt long. The sequences of the revertants showed the typical random removal of the inserts, leaving a stem with a 4- to 10-nt loop (not shown). Interestingly, in B.5 the complete insert except for 1 nt is deleted (Fig. 3).

**Evolution of the base-paired mutants C and D.** Evolution of mutants C and D with a base-paired insert is completely different from that of the unpaired mutants A and B. Three randomly picked plaques of each mutant were cycled in F<sup>+</sup> cells; no differences from the cDNA sequence were observed up to 10 passages (Fig. 6). This was not unexpected since both the C and D mutants have titers close to that of the wild type (Table 1); i.e., selection pressure is low. We conclude that extra information can be accommodated as long as it has the proper base pairing.

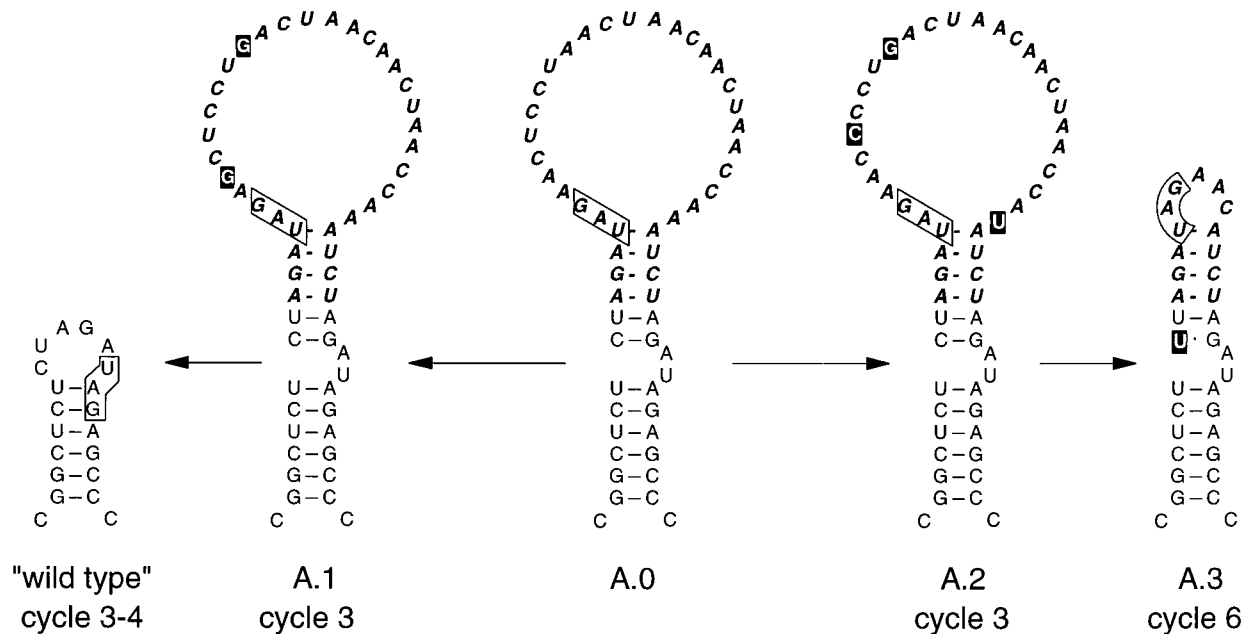


FIG. 7. Evolution of mutant A.0. Suppressor mutations are shown by a black squares. "wild type" signifies that the sequence has returned to wild type at the site of insertion but still contains the silent markers; i.e., it is a true revertant, not a contamination by real wild type. See the legend to Fig. 3 for further details.

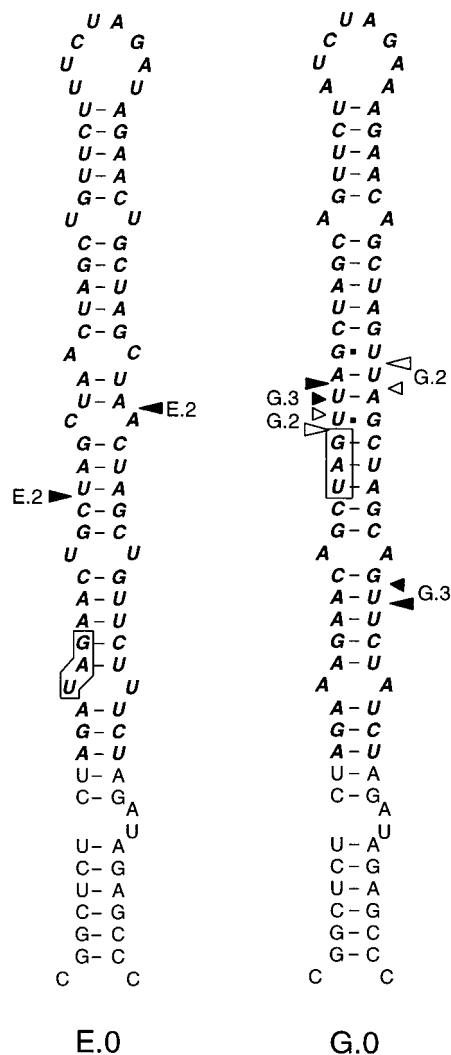


FIG. 8. Evolution of mutants E.0 and G.0. Arrowheads indicate the crossover points for E.2 and the possible crossover points for G.2 and G.3. For G.1, there are multiple possibilities (not drawn). E.1 cannot be drawn because of uncertainty about the sequence (see Fig. 9 for a linear representation). Revertants were sequenced after three cycles. See the legend to Fig. 3 for further details.

At cycle 11, mutant C had lost 1 nt from the insert, causing the stop codon of the A protein to move to the top of the helix (Fig. 6, C.1). Possibly, translation termination is unfavorable when the stop codon is halfway a stem structure (see Discussion for a possible explanation). Mutant D remained unaltered.

**Is there a limit to double-stranded RNA inserts?** To examine this point, we doubled the size of the inserts to 68 nt, using the same oligonucleotides as used for mutants C and D (see Materials and Methods). This yielded mutants E and G, in which the inserts are duplications of mutants C and D, respectively (Fig. 8). The infectivity of the E and G cDNA clones was at most 100-fold lower than that of the wild type (Table 1).

At cycle 3, revertants E.1 and E.2 lost about half of the original hairpin by random recombination (Fig. 8 and 9). For mutant G, there was a similar development. Of three descendants analyzed, one (G.1) lost the upper half of the hairpin by precise, and the other (G.2) did so by imprecise, homologous recombination, while G.3 was the result of apparently random

recombination. The evolution of mutants E and G shows that there is an upper limit to a structured extension of an existing hairpin.

## DISCUSSION

**RNA versus DNA recombination.** The insertions in MS2 were introduced via an infectious cDNA clone. In most cases, we do not recover those inserts at the viral RNA level because the extra sequence has already been (partially) lost before we can obtain the first RNA sequence (cycle 3; see Materials and Methods). This leaves the theoretical possibility that our revertants arise by DNA recombination (at the plasmid level). For several reasons, this seems unlikely. DNA recombination can occur between short repeats such as those created by the restriction sites present in our starting mutants (8). However, the *Xba*I and *Afl*III repeats are still present in the majority of the revertants, while the actual crossover points are very diverse and as a rule do not have sequence identity. Furthermore, several mutants (E and G) contain rather long repeats that are not exploited to yield the stable revertants mutants C and D, respectively (except G.1). Also, several revertants lost part of the insert during subsequent infection cycles, when we were past the cDNA stage. The evolution of A.2 to A.3 involving the removal of 21 nt between cycles 3 and 6 is a clear illustration (Fig. 7). Finally, we have observed that inserts in the 5' UTR are about 100 times more deleterious than inserts in the *Xba*I site at the end of the A-protein gene (Table 1). This is not easily explained by unequal DNA recombination frequencies, but the difference makes sense when one considers that inserts in the 5' UTR adversely affect replication and thus diminish the RNA pool that fuels recombination. We therefore believe that our revertants arose by RNA recombination driven by the MS2 replicase.

**Homologous versus random recombination.** All of our inserts are flanked by a direct repeat of at least 4 nt. Apparently, such a repeat is not sufficient to stimulate homologous recombination, since in the majority of our revertants we can detect neither a common motif nor a shared sequence at the crossover site (constructs L, M, P, Q, A, and B). The only common denominator in these revertants is the ability to form a base-paired structure. As evolutionary processes are not anticipatory, we assume that we are looking at the survivors of a recombinational process that appears completely random (but see below).

Only when the direct repeat approaches 10 to 12 bases, like in mutant A.0 and revertant A.1, do we score homologous recombination. This observation is consistent with previous findings for brome mosaic virus, in which case it was found that homologous, but not random, recombination drops dramatically upon a reduction of sequence homology from 15 to 9 nt (22). Nevertheless, the presence of even a 34-base repeat as in mutants E and G does not exclusively lead to homologous recombination, as shown by the appearance of revertants E.1, E.2, and G.3. The creation of these revertants may be induced by the presence of the long stem structure (see below).

**Recombination frequencies.** Some information about this parameter can be deduced from Table 1. Here, we can distinguish two kinds of infectious clones. In one kind, the emerging phages have the same sequence as the cDNA (R, C, and D), indicating that there is no strong selection pressure, while the slightly lower titer indicates a slightly lower fitness of these phages. In the other kind, the insertion is not recovered in the live virus probably because it is lethal. In these instances, a recombinational event (a deletion) seems required for viability, e.g., L and M. We suppose that the titers obtained for those

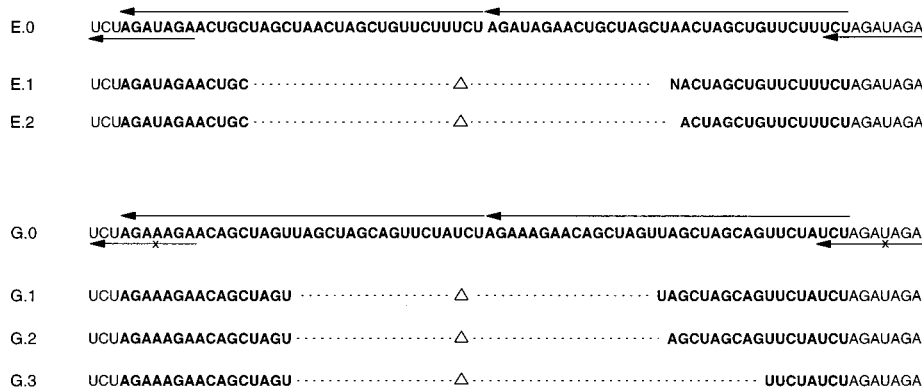


FIG. 9. Linear representation of the starting mutants E.0 and G.0 and of the emerging revertants. Arrows identify direct repeats that could favor homologous recombination. See the legend to Fig. 2 for further details. N is a nucleotide which we could not identify.

constructs reflect the recombination frequency. For instance, the probability to create a viable revertant from construct L is  $2 \times 10^7/10^{11} = 2 \times 10^{-4}$ . Assuming that every recombinant with a loop of 4 to 10 nt is viable, one can calculate that there are about 50 combinations to obtain such loops. This means that the probability of creating a specific single deletion is  $4 \times 10^{-6}$ . This number is substantially higher than found previously by Palasingam and Shaklee (26). These authors obtained a value of about  $10^{-8}$  for the chance of precisely deleting an 8-nt insert from a reading frame in Q $\beta$  RNA. These different frequencies suggest that a mechanism exists by which our large RNA-structure-disrupting inserts raise the random recombination frequency at the site of the mutation and nowhere else.

**Induced random recombination?** In vitro measurements of Axelrod et al. (1) with pure Q $\beta$  replicase and small self-replicating RNA molecules have shown that replication of structured RNA is more proficient than that of unstructured RNA. As their assay system did not contain any RNases, this difference could be attributed to unsuccessful strand separation. For mutants A and B (and possibly L, M, P, and Q), containing large unstructured loops, one could indeed envisage extensive duplex formation between plus and minus strands beginning at the loop sequences. This may pause or stall an approaching replicase, leading to drop-off and thus to increased recombination (by copy choice) (15) similar to what has been found for nonhomologous recombination in brome mosaic virus (21). Although pausing of the replicase is a process which normally accompanies the copying process (17, 27), we suggest that the extensive duplexes which may form in our mutants lead to much more pronounced stalling and to eventual drop-off, which will favor recombination.

Alternatively, RNase III may cut these long uninterrupted duplexes. Finally, the large unstructured loops may fall prey to endonucleases. Regardless of which RNase causes the cuts, the result will be the same: an increased drop-off of the replicase and thus a higher probability for a restart on the same or another RNA template.

Similarly, the long base-paired inserts of E and G may stall the replicase by their high stability. If stable stems or single-stranded inserts induce recombination, this means that the structure of wild-type viral RNA is optimized not to cause excessive stalling of the replicase by the presence of too much or too little base pairing. In this respect, it is noteworthy that, in fact, the secondary structure models of MS2 RNA and Q $\beta$  RNA show neither long single-stranded regions nor long uninterrupted stems (2, 29, 30).

**The position of the stop codon of the A-protein gene.** The assumption that single strandedness is harmful for the phage may also explain the fine tuning of mutant C. In this mutant, a single nucleotide was deleted upon continued evolution, causing the stop codon to move away from the wild-type position in the reading frame to a further downstream position coinciding with the top of the helix (Fig. 6). A ribosome terminating halfway along the stem in mutant C.0 covers the 5' side of the stem but leaves the 3' side unprotected to, e.g., RNase. This problem does not arise with stop codons located near the top of a hairpin. Similar arguments may explain the instability of mutants E and G. Some support for this idea is our observation that the stop codons of RNA phage genes tend to be located in or next to the loop of a small hairpin.

It is interesting that in mutant D, the stop codon of the A-protein gene is displaced three codons in the 3' direction compared with the wild type, yet mutant D containing an extended A protein is more stable than mutant C encoding the wild-type A protein.

**The flexibility of the C-terminal part of the A protein.** Finally, we wish to comment on the fact that most inserts in the XbaI site lead to an extension of the A protein. The great variety in size and composition of the C terminus of the A

TABLE 2. C termini of the A proteins in revertants

C-terminal amino acids	Revertant or wild type
LSRTILDRALNRSLKHGF .....	M.3
LSRVSLDRALNRSLKHGF .....	L.4
LSRVIDRALNRSLKHGF .....	L.3
LSRCLGVLSR .....	B.2
LSKIEPSTGV .....	B.5
LSDRTAS .....	C.1
LSRVGSR .....	L.1
LSRKNS .....	D.0, G.1, G.2, G.3
LSRTII .....	M.1
LSRTNP .....	M.2
LSRVSR .....	L.6
LSRCG .....	B.3
LSRFG .....	B.1
LSRTP .....	M.4
LSRVI .....	L.2
LSRVL .....	L.5
LSRF .....	B.4
LFR .....	A.3
LSR .....	Wild type, A.1, A.2, C.0, E.1, E.2
L .....	M.5



proteins in the revertants (Table 2) suggests that the selection pressure on this part of the protein is rather weak. In some revertants, the reading frames of the A-protein and coat protein genes overlap by 11 nt. Since the maturation gene is poorly expressed, this overlap probably has no consequences for the translation of the coat protein gene.

#### ACKNOWLEDGMENTS

We thank Herman Groeneveld for providing pHG227 and pHG229, and we thank Maarten de Smit for constructive comments during preparation of the manuscript.

#### REFERENCES

- Axelrod, V. D., E. Brown, C. Priano, and D. R. Mills. 1991. Coliphage Q $\beta$  RNA replication: RNA catalytic for single-strand release. *Virology* **184**: 595–608.
- Beekwilder, M. J., R. Nieuwenhuizen, and J. van Duin. 1995. Secondary structure model for the last two domains of single-stranded RNA phage Q $\beta$ . *J. Mol. Biol.* **247**:903–917.
- Biebricher, C. K., and R. Luce. 1992. *In vitro* recombination and terminal elongation of RNA by Q $\beta$  replicase. *EMBO J.* **11**:5129–5135.
- Chelladurai, B., H. Li, K. Zhang, and A. W. Nicholson. 1993. Mutational analysis of a ribonuclease III processing signal. *Biochemistry* **32**:7549–7558.
- Clarke, D. K., E. A. Duarte, S. F. Elena, A. Moya, E. Domingo, and J. Holland. 1994. The red queen reigns in the kingdom of RNA viruses. *Proc. Natl. Acad. Sci. USA* **91**:4821–4824.
- Court, D. 1993. RNA processing and degradation by RNase III, p. 71–116. *In* J. G. Belasco and G. Brawerman (ed.), *Control of messenger RNA stability*. Academic Press, New York.
- Domingo, E., E. Martínez-Salas, F. Sobrino, J. C. de la Torre, A. Portela, J. Ortín, C. López-Galíndez, P. Pérez-Breña, N. Villanueva, R. Nájera, S. VandePol, D. Steinhauer, N. de Polo, and J. J. Holland. 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. *Gene* **40**:1–8.
- Ehrlich, S. D. 1989. Illegitimate recombination in bacteria, p. 799–832. *In* D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
- Eigen, M., and C. K. Biebricher. 1988. Sequence space and quasispecies distribution, p. 211–245. *In* E. Domingo, J. J. Holland, and P. Ahlquist (ed.), *RNA genetics*, vol. 3. CRC Press Inc., Boca Raton, Fla.
- Groeneveld, H., K. Thimon, and J. van Duin. 1995. Translational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding? *RNA* **1**:79–88.
- Hjalt, T. Å. H., and E. G. H. Wagner. 1995. Bulged-out nucleotides protect an antisense RNA from RNase III cleavage. *Nucleic Acids Res.* **23**:571–579.
- Horiuchi, K. 1975. Genetic studies of RNA phages, p. 29–50. *In* N. D. Zinder (ed.), *RNA phages*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Jacobson, A. B., and M. Zuker. 1993. Structural analysis by energy dot plot of a large messenger RNA. *J. Mol. Biol.* **233**:261–269.
- Kozlovskaya, T. M., I. Cielens, D. Dreilina, A. Dislers, V. Baumanis, V. Ose, and P. Pumpens. 1993. Recombinant RNA phage Q $\beta$  capsid particles synthesized and self-assembled in *Escherichia coli*. *Gene* **137**:133–137.
- Lai, M. M. C. 1992. RNA recombination in animal and plant viruses. *Microbiol. Rev.* **56**:61–79.
- Miller, J. H., D. Ganem, P. Lu, and A. Schmitz. 1977. Genetic studies of the lac repressor. *J. Mol. Biol.* **109**:275–301.
- Mills, D. R., C. Dobkin, and F. R. Kramer. 1978. Template-determined, variable rate of RNA chain elongation. *Cell* **15**:514–550.
- Moody, M. D., J. L. Burg, R. DiFrancesco, D. Lovern, W. Stanick, J. Lin-Goerke, K. Madhavi, Y. Wu, and M. P. Farrell. 1994. Evolution of host cell RNA into efficient template RNA by Q $\beta$  replicase: the origin of RNA in untemplated reactions. *Biochemistry* **33**:13836–13847.
- Munishkin, A. V., L. A. Voronin, and A. B. Chetverin. 1988. An *in vivo* recombinant RNA capable of autocatalytic synthesis by Q $\beta$  replicase. *Nature (London)* **333**:473–475.
- Munishkin, A. V., L. A. Voronin, V. I. Ugarov, L. A. Bondareva, H. V. Chetverina, and A. B. Chetverin. 1991. Efficient templates for Q $\beta$  replicase are formed by recombination from heterologous sequences. *J. Mol. Biol.* **221**:463–372.
- Nagy, P. D., and J. J. Bujarski. 1993. Targeting the site of RNA-RNA recombination in brome mosaic virus with antisense sequences. *Proc. Natl. Acad. Sci. USA* **90**:6390–6394.
- Nagy, P. D., and J. J. Bujarski. 1995. Efficient system of homologous RNA recombination in brome mosaic virus: sequence and structure requirements and accuracy of crossovers. *J. Virol.* **69**:131–140.
- Olsthoorn, R. C. L., G. Garde, T. Dayhuff, J. F. Atkins, and J. van Duin. 1995. Nucleotide sequence of a single-stranded RNA phage from *Pseudomonas aeruginosa*; kinship to coliphages and conservation of regulatory RNA structures. *Virology* **206**:611–625.
- Olsthoorn, R. C. L., N. Liciis, and J. van Duin. 1994. Leeway and constraints in the forced evolution of a regulatory RNA helix. *EMBO J.* **13**:2660–2668.
- Olsthoorn, R. C. L., S. Zoog, and J. van Duin. 1995. Coevolution of helix stability and Shine-Dalgarno complementarity in a translational start region. *Mol. Microbiol.* **15**:333–339.
- Palasingam, K., and P. N. Shakklee. 1992. Reversion of Q $\beta$  RNA phage mutants by homologous RNA recombination. *J. Virol.* **66**:2435–2442.
- Priano, C., F. R. Kramer, and D. R. Mills. 1987. Evolution of the RNA coliphages: the role of secondary structure during RNA replication. *Cold Spring Harbor Symp. Quant. Biol.* **52**:321–330.
- Remaut, E., P. Stanssens, and W. Fiers. 1981. Plasmid vectors for high-efficiency expression controlled by the *p<sub>L</sub>* promoter of coliphage lambda. *Gene* **15**:81–93.
- Skripkin, E. A., M. R. Adhin, M. H. de Smit, and J. van Duin. 1990. Secondary structure of the central region of bacteriophage MS2 RNA. Conservation and biological significance. *J. Mol. Biol.* **211**:447–463.
- Skripkin, E. A., and A. B. Jacobson. 1993. A two-dimensional model at the nucleotide level for the central hairpin of coliphage Q $\beta$  RNA. *J. Mol. Biol.* **233**:245–260.
- van Duin, J. 1988. Single-stranded RNA bacteriophages, p. 117–167. *In* H. Fraenkel-Conrat and R. Wagner (ed.), *The viruses—1988*. Plenum Press, New York.