Nucleotide sequence of a single-stranded RNA phage from Pseudomonas aeruginosa: kinship to coliphages and conservation of regulatory RNA structures

Olsthoorn, R.R.C.L.; Garde, G.; Dayhuff, T.; Atkins, J.F.; Duin, J. van

VIROLOGY 206, 611–625 (1995)

Nucleotide Sequence of a Single-Stranded RNA Phage from *Pseudomonas aeruginosa*: Kinship to Coliphages and Conservation of Regulatory RNA Structures

R. C. L. OLSTHOORN,*,[1] G. GARDE,†,[1] T. DAYHUFF,‡,§ J. F. ATKINS,†,‡ AND J. VAN DUIN*,[2]

*Department of Biochemistry, Leiden Institute of Chemistry (LIC), Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands; †Department of Biochemistry, University College, Cork, Ireland; and ‡Howard Hughes Medical Institute and §Department of Human Genetics, 6160 Eccles Genetics Building, University of Utah, Salt Lake City, Utah 84112

Received May 24, 1994; accepted October 10, 1994

We report the complete nucleotide sequence of the single-stranded RNA phage PP7 from *Pseudomonas aeruginosa*. There are three open reading frames which code for apparent protein homologues of the single-stranded RNA coliphages, i.e., maturation protein, coat protein, and replicase. A fourth overlapping reading frame exists that probably encodes a lysis protein, similar to what has been found in the group A coliphages such as MS2. The genetic map of PP7 is colinear with group A coliphages and we accordingly classify the phage as a levivirus. There is, generally speaking, no significant nucleotide sequence identity between PP7 and the coliphages except for a few regions where homologous parts of proteins are encoded, most notable in the replicase gene. In these regions the nucleotide sequence similarity between PP7 and MS2 is no greater than between PP7 and the group B coliphages such as Q_β. Surprisingly, Q_β and MS2 are no closer to each other than they are to PP7. Several regulatory RNA secondary structure features that are present in the coliphages were identified also in PP7 RNA although the sequences involved cannot be aligned. Among these are the coat protein binding helix at the start of the replicase gene, structures at the 5′ and 3′ terminus of the RNA, a replicase binding site, and the structure of the coat protein cistron start. Some of these features resemble MS2 type coliphages but others the Q_β type. These findings suggest that PP7 is related to the coliphages but branched off before the coliphages diverged into separate groups.

INTRODUCTION

The single-stranded RNA bacteriophages infect a wide range of Gram-negative bacteria, provided they express the proper F- or polar pili on their surface. Initially, these phages were discovered in *Escherichia coli* by Loeb and Zinder, but soon after similar creatures were found in Caulobacter and Pseudomonas (for a review see Zinder, 1975).

Based on different physical and serological properties of a large number of isolates from different ecological niches in numerous countries, the single-stranded RNA coliphages have been divided into four groups (I to IV). Groups I and II with MS2 and GA as the type species show relatively small differences among each other and are collectively called group A. Similarly, groups III and IV with Q_β and SP as best known representatives are similar and together form group B (Furuse, 1987). Genetic maps of MS2 and Q_β are shown in Fig. 1. Their basic difference is the presence of a readthrough coat protein in group B, which gets incorporated into the virion and is necessary for infection. Group A phages, on the other hand, have a separate gene for the lysis function while in group B the maturation protein can cause cell lysis (for a review see van Duin, 1988).

The RNA phages of Caulobacter and Pseudomonas have not been examined in much detail. They have the same size and icosahedral structure as their *E. coli* counterparts. The *Pseudomonas aeruginosa* phage PP7 was isolated by Bradley (1966). While the coliphages adsorb to F pili (Crawford and Gesteland, 1964), PP7 is not male specific and does not attach to the Pseudomonas sex factor FP pilus, but instead adsorbs on the sides of polar pili (Bradley, 1972), which are crucial for virulence. The diameter of these pili is only 5.2 nm compared to 9 nm for the *E. coli* F pilus (review Paranchych and Frost, 1988). The other RNA phage isolated from Pseudomonas, PRR1, adsorbs to different pili. PRR1 infects a wide range of Gram-negative bacteria harboring the promiscuous P-type drug resistance plasmid by adsorbing to the plasmid-encoded pilus. The amino acid sequences of the coat proteins of PRR1 and PP7 were determined and they showed relatedness to those of MS2 and Q_β (Dhaese *et al.*, 1979, 1980; Golmohammadi *et al.*, 1993).

In contrast, the coliphages have been the subject of extensive studies. In particular, the availability of Q_β RNA replicase in pure, well-characterized form has allowed a thorough *in vitro* analysis of kinetics, template require-

[1] The first two authors participated equally in this study and should be considered joint first authors.

[2] To whom reprint requests should be addressed.

0042-6822/95 $6.00

ment, specificity, and binding sites of the enzyme (Biebricher and Luce, 1993 and references therein; Axelrod *et al.,* 1991; Barrera *et al.,* 1993).

MS2 RNA has served as a model system to unravel translational control mechanisms (Berkhout *et al.,* 1987; Adhin and van Duin, 1990; van Himbergen *et al.,* 1993). It was a surprise to find that most of this regulation could be explained by the interplay of ribosome movement and the dissolution and reformation of base-paired regions. Clearly, if one is to understand the molecular biology of RNA phages, and of RNA viruses in general, one needs to know the RNA secondary structure. Accordingly, we and others have begun to expand the structural analysis to regions not directly implicated in ribosome binding (Skripkin *et al.,* 1990; Adhin *et al.,* 1990b; Skripkin and Jacobson, 1993).

The sequences of five coliphages have been published (Fiers *et al.,* 1976; Mekler, 1981; Inokuchi *et al.,* 1986, 1988; Adhin *et al.,* 1990a) and the sequence of six others has been determined in Leiden but not yet reported. Based on such information phylogenetically supported models for the coliphages are emerging and it becomes desirable to have a reference frame to compare the *E. coli* phages with. This is one reason why we determined the complete sequence of PP7. We wanted to know whether features of the structure conserved among the coliphages are also maintained in the Pseudomonas phages, i.e., is there an archaetype structure for all RNA phages? Furthermore, since PP7 lives in a host with a high G+C content in its chromosome, one could ask if this nucleotide bias is reflected in the resident RNA phages.

A second reason for investigating PP7 was that it has an "extra" protein in its virion. Dhaese (1979) prepared a fingerprint of this protein, p25, which suggested that it has the same amino terminal sequence as the coat protein which has a molecular weight of 13,874 Da. It could not be a coat read-through protein analogous to that of group B phages since PP7 RNA has no space for an in-frame read-through product even if it were only a moderate fraction of the size found in Q_β. Because of our *in vitro* study on MS2 translation products which arose by frameshifting near the end of the coat (and replicase) gene(s), it was attractive to think that PP7 may package an equivalent protein.

MATERIALS AND METHODS

Virus and RNA isolation

PP7 phage was obtained from P. Dhaese and M. van Montagu (University of Ghent, Belgium) and its host *P. aeruginosa* was a gift from B. Holloway (Monash University, Australia). *P. aeruginosa* PA01 (Holloway, 1969) was infected with a fresh lysate of PP7 phage at conditions similar to amplification of coliphages (Adhin *et al.,* 1990a). After cell lysis the lysate was purified by ammonium sulfate precipitation and CsCl centrifugation as described (Voorma *et al.,* 1971), except that freon extraction was omitted. PP7 RNA was isolated by two successive phenol extractions followed by alcohol precipitation. The RNA migrated as a single band and slightly faster than MS2 RNA during electrophoresis in a 1% agarose gel.

Synthesis and cloning of cDNA

Regions 1–3300. cDNA clones covering the regions 1–960, 390–2570, 1900–3000, and 2240–3300 were obtained as follows. 12 pmol of PP7 RNA and 30 pmol of random hexamers were heated for 1 min at 42° in a 20-μl reaction mixture containing 20 m*M* DTE, 0.5 m*M* dNTP, 50 m*M* Tris–HCl (pH 8.3), 40 m*M* KCl, and 8 m*M* $MgCl_2$. First strand synthesis was started by addition of 16 units M-MuLV reverse transcriptase (Pharmacia). After 10 min at room temperature the reaction was continued at 30° for an additional 10 min and concluded for 3.5 hr at 42°. The RNA–cDNA hybrid was purified by two phenol/chloroform extractions, passage through a 1-ml Sephadex G25-column, and ethanol precipitation. The RNA–cDNA hybrid was tailed with dCTP by terminal transferase (Bethesda Research Laboratories) and subsequently annealed to an oligo(dG)-tailed pUC9 vector (Pharmacia) as described earlier (Adhin *et al.,* 1990a). JM101 was transformed with the annealing mixture and clones with inserts were selected for further analysis by restriction mapping followed by subcloning into M13mp18 and M13mp19.

Region 3216–3588 (3′ end). The tight RNA secondary structure at the 3′ end of several coliphages (Adhin *et al.,* 1990a) can make cloning difficult. We proceeded as follows. RNA (30 pmol) in 10 μl distilled H_2O was heated at 60° for 5 min and then rapidly chilled in ice. Poly(A) buffer (final concentration 50 m*M* Tris–HCl (pH 7.9), 250 m*M* NaCl, 10 m*M* $MgCl_2$, 6.5 m*M* $MnCl_2$, 50 μg bovine serum albumin), ATP (125 μM), 2 units RNasin (Promega), and 5 units poly(A) polymerase (BRL) were added and the mixture incubated at 37° for 30 min. The A-tailed RNA was purified as described for the RNA–cDNA hybrid. cDNA synthesis was performed as described above except that 18 units AMV reverse transcriptase (Promega) were used and first strand synthesis was initiated with either oligo DUI416 (5′ dTTTTTG(G/C)AT(C/G)CTTTTT-TTTTTTGGG 3′) or oligo DUI570 (5′ dT_{20}GG 3′). DUI416 was specially designed for cloning 3′ ends of coliphages as they terminate with the conserved -CCCA sequence. However, attempts to clone the 3′ end of PP7 with this primer were unsuccessful since DUI416 tended to hybridize with a stretch of C-residues around position 3353. Therefore we used a primer with two Gs at the 3′ end (DUI570). It turned out that PP7 RNA ends with -ACCA, which explained our failures with DUI416.

The cDNA obtained with DUI570 was amplified by PCR using a second primer (DUI503) homologous to bases 3216 to 3241 and 1.5 units Taq polymerase (Promega). The resulting PCR product was tailed with dCTP and cloned into pUC9 as described above. These experiments produced four clones covering the remaining 3′ end. For sequencing purposes some clones were shortened by deleting a *Sal*I-fragment. A *Sal*I-site is present in PP7 at position 3334 and also in pUC9.

Sequencing

The regions 1–450 and 2240–3300 were sequenced as follows. Restriction fragments of 200–300 bp were subcloned into M13 vectors and sequenced in both directions. Clones covering the regions 390–2570 and 3216–3588 were directly sequenced by the supercoil sequencing procedure (Chen and Seeburg, 1985). The sequence of both strands was determined using plus and minus strand-specific primers. DNA sequencing was carried out by the chain termination method (Sanger *et al.*, 1977) using the T7 sequencing kit (Pharmacia).

The number of G-residues at the 5′ end of PP7 could not be determined because of the G-tail derived from the cDNA cloning procedure. For this reason reverse transcription of the RNA using a primer complementary to bases 43–62 (DUI600) was applied to solve this problem. The conditions for the reverse transcription were as described earlier (Skripkin *et al.*, 1990) except that the dNTP concentration was 10 times higher and the chase reaction and ethanol precipitation were omitted.

Nucleotides 1182, 1360, and 2050 could not be identified by DNA sequencing. The identity of position 1360 was crucial for determining the stop-codon of the maturation gene. By reverse transcription of the RNA with a primer complementary to 1444–1466 (DUI812) two bases were found. An identical result was obtained when the reaction was performed with dITP instead of dGTP to minimize band compression in the gel.

Nucleotide 1182 was clarified by sequencing the first strand of cDNA obtained in the previously mentioned reaction with DUI 812. The cDNA was treated with RNAse to remove the RNA template and after phenol/choroform extraction the cDNA was purified over a G25 column to remove all nucleotides which could interfere with the sequencing. Sequencing was carried out with AMV reverse transcriptase at 55° and a primer DUI 811 homologous to 1140–1165 under the same conditions as for reverse transcription of the RNA. For nucleotide 2050 all the afore mentioned methods were tried but failed; since *Ava*II did not cut here we know it must be either a G, A, or U (symbol "D"). The complete sequence has been deposited with GenBank, Accession No. X80191.

Software

Computer programs were from the software package of the Genetics Computer Group of the University of Wisconsin, release 7.3 (Devereux *et al.*, 1984). Alignments were carried out with the PILEUP program. Homology percentages were determined with the GAP program which uses the Needleman and Wunsch algorithm. Secondary structure predictions were made by the FOLD program (Zuker and Stiegler, 1981) and by eye.

RESULTS AND DISCUSSION

Genome organization

PP7 RNA is 3588 nucleotides long; this size is similar to the group I phages such as MS2 and fr which are 3569 and 3575 nucleotides long, respectively. There are at least three, and probably four, open reading frames in the sequence and their positions are given in Fig. 1. Based on protein homology with the coliphages (see below), the size of the maturation protein from virions (not shown), and the products of cell free protein synthesis (Davies and Benike, 1974), we have designated these genes as maturation, coat, and replicase. We also propose that a small overlapping reading frame encodes the lysis function.

The map resembles that of the group I coliphages. One difference is the 100 bases longer maturation protein gene, which results partly from a displaced stop codon overlapping the start of the coat protein gene. We do not think the overlap has any regulatory consequences. The maturation gene is translated with very low frequency and coat protein expression is independent of it in the coliphages. Furthermore, we have made several mutant MS2 phages in which the maturation reading frame was extended to overlap that of the coat gene by some 10 nucleotides. Such mutants were viable and they had a titer that was close to that of wild type (Olsthoorn and van Duin, unpublished observations).

Protein sequences

Maturation protein. The maturation protein probably occurs at one copy per virion and has two known functions. It matures the virus particle by preventing the RNA from dangling out of the virion, thereby conferring protection against RNase, and it is necessary for infection. Upon contact with the host's pili the maturation protein is cleaved leading to release of the RNA (Zinder, 1975). In Q_β and presumably in other group B phages the maturation protein also possesses a lysis function (Karnik and Billeter, 1983; Winter and Gold, 1983).

Although unlikely (see below), we cannot exclude the possibility that the PP7 maturation protein causes cell lysis. Its homology to the group B and group A counterparts is equally low (~20%), which is not surprising since PP7 adsorbs to different pili. Nevertheless, when the proteins are aligned three stretches of significant homology are found (boxed in Fig. 2). The C-terminal one was pre-

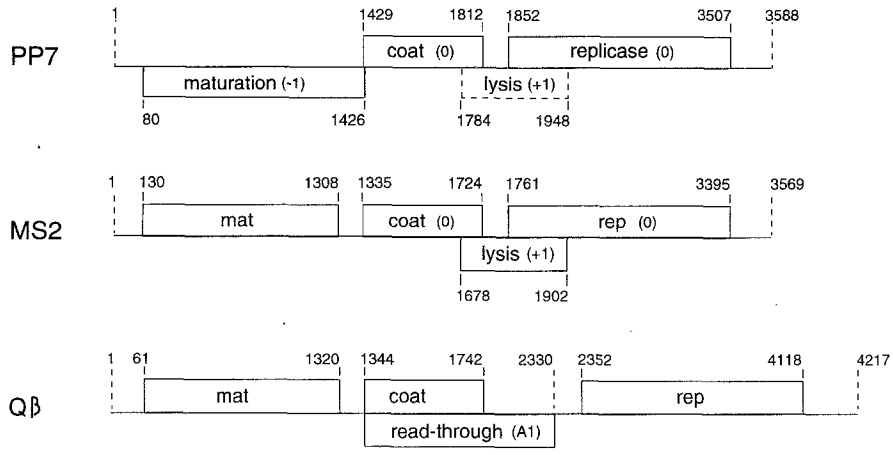

FIG. 1. Genetic map of the single-stranded RNA phage PP7 from *Pseudomonas aeruginosa.* Where necessary, relative reading frames are indicated. For comparison the maps of the coliphages MS2 (group I) and Q_β (group III) are given.

viously identified by Mekler (1981) as homologous between MS2 and Q_β. The other two domains, which are rich in aromatic residues, have not been identified before and have no counterparts in other proteins, as revealed by a FASTA search of the Swiss protein bank.

Coat protein. The coat proteins of all single-stranded RNA phages are remarkably uniform in size (Fig. 3). With 127 amino acids (methionine is cleaved off) PP7 contains the shortest version so far. Our sequence differs at five contiguous positions from the one published by Dhaese *et al.* (1980). In that report, residues 111–115 are mentioned as VQATS. This seems a scrambled version of the RNA predicted sequence ATSQV. We believe that the RNA-derived sequence is correct. When homologies are scored we note that the PP7 coat protein is about equally distant from group A and B counterparts with a slightly better fit with the group A proteins (Table 1). At the nucleotide level PP7 is equidistant to all four coliphages in the coat gene (Table 1). Still, A and B coliphages show somewhat better protein homology with each other than with PP7. A recent evaluation of the coat proteins of RNA phages has been published by Golmohammadi *et al.* (1993).

p25 protein. Tryptic fingerprint analysis showed that p25 and the coat protein are related (Dhaese, 1979). Using p25 that was gel purified three times, we found that it has the same N-terminal amino acid sequence as coat protein. As judged by SDS–PAGE, p25 has an apparent molecular weight 12 kDa larger than coat protein. From the RNA sequence, readthrough of the coat protein gene terminator would give an extension of 8 amino acids, whereas a −1 or +1 frameshift just prior to the coat gene terminator would give extensions of 16 or 45 amino acids, respectively. Either of the putative frameshift proteins upon tryptic digestion would yield C-terminal peptides with molecular weights of 3–4 kDa absent from digested coat protein, but no such peptides were found (data not shown). Analysis with the protease Asp-N, which cleaves at the amino terminal side of aspartate residues, also yielded data inconsistent with a frameshift model for p25 (Garde and Dayhuff, unpublished observations).

Different PP7 phage preparations, grown and purified through two cesium chloride gradients, were variably pigmented red, despite the use of the same inoculum. When samples of these preparations were fractionated on SDS–PAGE, they showed differing amounts of the minor virion protein p25, varying from 1 to 10 molecules per virion (as measured against the maturation protein, which is present in coliphages at one molecule per virion). The phage preparations with the most p25 corresponded to those with the deepest red color, implicating p25 as the source of red pigment. Furthermore, when enough purified phage is loaded on SDS–PAGE, the p25 protein is visible as a dark red or brown band as it migrates through the gel (data not shown). Although small bits of identity exist between the cysteine domain of PP7 coat protein and that of some nonheme iron–sulfur proteins, spectroscopic analysis reveals that any possible contribution of iron–sulfur bridges occurs at less than a few percent (Dayhuff, unpublished observations). Our current explanation for p25 is a nondirected collision of heme with coat protein leading to a thioester linkage to cysteine stabilized by formation of a second thioester linkage to a nearby cysteine (cysteine is at positions 67 and 72). When this hemocoat protein is assembled into a growing virion one of the thioester linkages shifts from an intramolecular bond to an intermolecular bond creating a coat dimer. This model is supported by substantial but incomplete evidence (Dayhuff, unpublished observations).

Previously it was noted that MS2 phage with defined coat protein mutants yielded an extra band on SDS–PAGE which was most likely due to a small proportion

```
SP     1 ..................................MPTLPRGLRFGSNGEVLNDFEALWFPERHTVDLSNGT..CKLTGY
Qß     1 ..................................MPKLPRGLRFGADNEILNDFQELWFPD.LFIESSDTHPWYTLKGR
GA     1 ....................MFPKSNIDRNYKVKLISYDKKGKLVSDDS.FEQVENYLFQNRSTTYKPGYIR......R
MS2    1 ....................VRAFSTLDRENE....TFVPSVRVYADGE.TED.NSFSLKYRS.NWTPGRFNSTGAKTK
PP7    1 MLEEHYYSRKTCGSGGGLSVQRTGYPQTDPPVFKAVNSFERAPLIVGD.....HVHSLPYWSRKITLECTPYPFT.....

SP    44 ITNLPGYSDIFPNKGVTAA.RTPYRSTVPVNHLGYRPVTTVEYIPDGTYVRLDGHVKFE.GDLVNGSVDLTNFVISLAAQ
Qß    45 VLN.AHLDDRLPNVGGRQVRRTPHRVTVPIASSGLRPVTTVQYDPAALSFLLNARVDWDFGNGDSANLVINDFLFRTFAP
GA    53 DFRRPT.NF.WNGYRCFNQPVGTFTRKLSDGGR...QVADYGIVNPNKFTANSQHLGDNMVI.YPGPFSIN..IDQRASV
MS2   53 QWHYPS.PYSRGALSVTSIDQGAYKRSGSSWGRPYEEKAGFGFSLDARSCYSLFPVSQNLTY.IEVP..QN..VANRAST
PP7   71 .LRWEG.GVHKDGIP.........DKVPHPSCDPGSELVSYTHSREVDCERFIGPVPSSFTD.PKVTSAQWKFLEDMADM

SP   122 GGFDYQSVIGPRFSARFSAFSTKYGVLLGEGRETLKYLLLVVRRMREGYRAVRRGDLKRLRNVISTFEPSTIKGKRARAE
Qß   124 KEFDFSNSLVPRYTQAFSAFNAKYGTMIGEGLETIKYLGLLLRRLREGYRAVKRGDLRALRRVIQSYH.............
GA   125 EVL............NKLSQSNLNIGVAIAEAKMTASLLAKQSIALIRAYTAAKRGNWREVLSQLL...............
MS2  127 EVL............QKVTQGNFNLGVALAEARSTASQLATQTIALVKAYTAARRGNWRQALRYLA...............
PP7  139 KAL............ADLNKSLVNLPMLYKERRETLKMVGNRLGGLVRVAHAAQD...RDLKRYFK...............

SP   202 FSQTYRDKLTGNKVEVRPSEGKWNSSSASDL WLEFRYGLMPLFYDI QSVMEDFMRVHKKIAKIQRFSAGHGKLETVS.SR
Qß   192 ...................NGKWKPATAGNL WLEFRYGLMPLFYDI RDVMLDWQNRHDKIQRLLRFSVGHGEDYVVEFDN
GA   179 ................ISEHRFRAPAKDLGGR WLELQYGWLPLMSDL KAA..YDLLTQTKLPAFMPLRVTRTVGGTHNYKV
MS2  181 ................LNEDR.KFRSKHVAGR WLELQFGWLPLMSDI QGA..YEMLTKVHLQEFLPMRAVRQVGTNIKLDG
PP7  190 ................ARRKDRRKVAEEVANG YLELIFGWLPLIGEL EGAIEYAELPDLDFIRCHGLHTLVLQSTPWDNSV

SP   281 FYPDV.HFSLEVTAVLQRRHR.WGVIYQDTGSFATFNNGRLVPVKDWKTAAFALLNPAEVA WEVTPYSFVVDWFVN VGDM
Qß   253 LYPAVAYFKLKGEITLERRHR.HGISYANREGYAVFDNGSLRPVSDWKELATAFINPHEVA WELTPYSFVVDWFLN VGDI
GA   242 RNVESAGDTWSYRHRLSVNYRIWYFISDP.......RLAWASSLG........LLNPLEIY WEKTPWSFVVDWFLP VGNL
MS2  243 R.LSYPAANFQTTCNISRRIVIWFYINDA.......RLAWLSSLG.......ILNPLGIV WEKVPFSFVVDWLLP VGNM
PP7  255 DVRSYPNWERAAGTRITGSVRTRGVVESRASVRTALRFNLETSLAG..DARRLGFEPISTT YDMIPLSFVVGWFSN FDKY

SP   359 LEQMGQLYRHVDVVDGFDRKDIKLKSVSVR..........VLTNDVAHVASFQLRQAKLLHSYYSRVHTVAFPQISPQL
Qß   332 LAQQGQLYHNIDIVDGFDRRDIRLKSFTIK..........GERNGRPVNVSASLSAVDL...FYSRLHTSNLPFATLDL
GA   307 IEAMSNPL.GLDIISGTKTWQLESKLNA.......TLPA.SGWS.........GTAKLTAYAKAYDRSTFYSFPTPLPYV
MS2  307 LEGLTAPV.GCSYMSGTVTDVITGESII.......SVDAPYGWTVERQ.....GTAK..AQISAMHRGVQSVWPTTGAYV
PP7  333 VRTLAPLI.GVTFETGSQNRRTTCELVGCTRFYPRTVSPPSGWFARWKDFPDGSLSEVSGLRRTDIRSVLSTLPDPDVRF

SP   428 DTEIRSVKHVIDS IALLTQRVKR ..............  450
Qß   398 DTTFSSFKHVLDS IFLLTQRVKR ..............  420
GA   369 KSPLSGL.HLANA LALINQRLKR ..............  390
MS2  372 KSPFSMV.HTLDA LALIRQRLSR ..............  393
PP7  412 HADVGLF.EISAG ISLLAQRYLK PLQRLLKRKSFFYGRT 449
```

FIG. 2. Amino acid alignment of the maturation proteins of phage PP7 and one representative of each of the four coliphage groups. SP and GA belong to group IV and II, respectively. The three regions with significant homology are boxed. Residues conserved in all phages are shown in bold face.

of stable coat protein dimers (see Fig. 6W in Peabody, 1989).

Lysis protein. The putative lysis (L) gene is in the same place as its MS2 counterpart, i.e., starting near the end of the coat protein gene in the +1 frame with respect to the coat gene, extending through the intercistronic region, and overlapping the 5′ region of the replicase gene. Like its MS2 counterpart (Atkins *et al.*, 1979b; Beremand and Blumenthal, 1979; Coleman *et al.*, 1983), the putative PP7 L protein has a positively charged N-terminal part followed by 30 mainly hydrophobic residues at the C-terminus. This composition is characteristic of the L proteins of bacteriophages. The hydrophobic domain is thought to cross the cytoplasmic membrane and to short-circuit its electric potential (reviewed recently by Young, 1993). The interaction with the membrane can apparently be achieved by a variety of hydrophobic amino acids since strong homologies have not turned up when comparing the sequences of all known RNA phage L proteins (our unpublished observations). Although the PP7 lysis protein with its 54 amino acids seems small compared to the 75 amino acids for the MS2 counterpart, it is known that truncated versions of the MS2 L protein are still functional as long as they retain their C-terminal 30 residues (Berkhout *et al.*, 1985; Goessens *et al.*, 1988).

The start codon of the PP7 L gene is UUG. This is an

```
 SP    1 .AKLNQVTLSKIGKNGDQTLTLTPRGVNPTNGVA.SLSEAGAVPALEKRVTVSVAQPSRNRKNFKVQIKLQNPTACT...
 Qß    1 .AKLETVTLGNIGKDGKQTLVLNPRGVNPTNGVA.SLSQAGAVPALEKRVTVSVSQPSRNRKNYKVQVKIQNPTACT...
 GA    1 .ATLRSFVLVDNGGTGNVTVV....PVSNANGVAEWLSNNSRSQAY..RVTASYRASGADKRKY..AIKLEVPKIVT.QV
MS2    1 ASNFTQFVLVDNGGTGDVTVA....PSNFANGVAEWISSNSRSQAY..KVTCSVRQSSAQNRKY..TIKVEVPKVAT.QT
PP7    1 ...SKTIVLSV....GEATRTLTEIQSTADRQIFEE.KVGPLVGRL..RLTASLRQNGA.KTAYRVNLKLDQADVVDCST

 SP   76 .RDACDPSVTRSAFADVTLSFTSYSTDEERALIRTELAALL..ADPLIVDAIDNLNPAY. 131
 Qß   76 ANGSCDPSVTRQAYADVTFSFTQYSTDEERAFVRTELAALL..ASPLLIDAIDQLNPAY. 132
 GA   71 VNGVELPGSAWKAYASIDLTIPIFAATDDVTVISKSLAGLFKVGNPIA.EAISSQSGFYA 129
MS2   71 VGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP.SAIAANSGIY. 128
PP7   70 SVCGELPKVRYTQVWSHDVTIVANSTEASRKSLYDLTKSLVATSQVEDLVVNLVPLGR.. 127
```

FIG. 3. Amino acid alignment of the coat proteins of PP7 and four representative coliphages. Residues conserved in all phages are shown in bold face.

unusual initiator triplet; however, this codon is also used for the L genes in phage fr and KU1, a recently sequenced group II phage. In this last phage the L protein is also short, i.e., 64 amino acids (Groeneveld and van Duin, submitted for publication).

Replicase. Extensive studies with Q_β indicated that the phage-encoded component of the RNA replicase forms a complex with EF-Tu and EF-Ts to build the active enzyme and that two additional host factors are required to produce the minus strand from the plus strand. These are ribosomal protein S1 and the product of the recently cloned *hfq* gene (HF1) (see Kajitani *et al.,* 1994). Despite its interaction with Pseudomonas rather than *E. coli* components, the replicase is the most conserved of the phage proteins. In particular, there is a large region in the center of the protein with extensive homology. In Fig. 4 this part is shown in brackets and conserved residues are in bold face. The degree of similarity in this region between bacteriophages is presented in Table 2 and basically shows the equidistance between PP7 and the various coliphages.

The YGDD motif, which is so far universally conserved among all RNA-dependent RNA replicases, occurs at position 338 (Kamer and Argos, 1984). Other strongly conserved residues between the coliphages and PP7 are the IDLN (1), DLS.ASD (2), MGNG (3), and FRESCG (4) motifs at position 230, 255, 304, and 374, respectively. Motifs 2, 3, and 4 have also been identified in other RNA-dependent RNA polymerases (RDRPS), although in somewhat different forms (Poch *et al.,* 1989; Bruenn, 1991). Motif 1, which is also present in the recently sequenced coliphages MX1, M11, and NL95 (Beekwilder, Nieuwenhuizen, and van Duin, submitted for publication) has not been found in other RDRPS. We do find strong homology with a region just upstream of the YGDD motif in polio-, coxsackie-, and human rhinoviruses which are closely related to the leviviruses (Bruenn, 1991).

TABLE 1

PERCENTAGE AMINO ACID IDENTITY BETWEEN THE COAT PROTEINS

	MS2	GA	Q_β	SP
PP7	22 (38)	23 (39)	19 (38)	21 (39)
MS2	—	60 (62)	19 (38)	24 (42)
GA	—	—	26 (43)	25 (41)
Q_β	—	—	—	80 (71)

Note. The identity at the nucleotide level is indicated in parentheses.

Conserved RNA secondary structures

The stable 5′ terminal hairpin. All coliphages have a stable stem structure closed by a tetraloop at their 5′ ends. PP7 RNA has the same feature as is shown in Fig. 5, where its 5′ terminal hairpin is compared to that of MS2, Q_β, and M11. The upper part of the stem always contains straight C·G pairs. The lower half is generally weaker by the presence of a bulge, a mismatch, G·U pairs, or a combination thereof.

We speculate that this hairpin at the terminus is important to ensure that the emerging nascent plus strand folds back on itself and does not anneal with the template strand. In this context it is probably significant that the loop sequence of the 5′ terminal hairpin in both MS2 and M11 is the complement of the tetraloops 5′GAAA and 5′UACG, respectively, which are known to confer increased hairpin stabilities (Heus and Pardi, 1991; Tuerk *et al.,* 1988). (The tetraloop sequences in PP7 and Q_β have not been examined in this respect.) This design then will promote intramolecular folding of the minus strand. The lower stability at the bottom part of the stem is believed to facilitate access of the replicase when plus strand synthesis has to start. Note that the G·U pairs will be C·A mismatches in the minus strand. Interestingly, similar, rather stable, hairpins have been found at the ends of artificial Q_β replicase templates (see also Biebricher and Luce, 1993).

The strong 5′ terminal hairpin makes it difficult to clone and sequence this part of the chain. Reverse transcriptase shows a strong stop some five nucleotides before the loop and transcripts through the loop show anomalous mobilities even under strong denaturing conditions.

Coat gene start. This part of the phage has been well studied in MS2 where the RNA folding has been

```
SP     1 MPKTASRRREITQLL...GKVDINFEDDIHMSIANDLFEAYGIPKLDSAEECINTAFPSLDQGVDTFR.VEYL.....RA
Qß     1 MSKTASSRNSLSAQLRRAANTRIEVEGNLALSIANDLLLAYGQSPFNSEAECISFS.PRFDGTPDDFR.INYL.....KA
GA     1 MFRFREIEKTLCMDRTR......DCAVRFHVYLQ.....................SLDLGSSDPLSPDFDGLAYL.....RD
MS2    1 MSKTTKKFNSLCIDLPR......DLSLEIYQSIA.....................SVATGSGDPHSDDFTAIAYL.....RD
PP7    1 ............METLSSAQLRDVVSSELGLPVPTQLPHR...............ELGGVEKSPEQFAKEYFLYNLLRKV

SP    72 EILSKFDGHPLGIDTE...AAAWEKFLAAEEGCRQTNERLSLVKYHDNSILSWGERVIHT.ARRKILKLIGESVPFGDVA
Qß    74 EIMSKYDDFSLGIDTE...AVAWEKFLAAEAECALTNARLYRPDYSEDFNFSLGESCIHM.ARRKIAKLIGDVPSVEGML
GA    51 ECLTKHPSLGDSNSDARRKELAYAKLMDSDQRCKIQNSNGY..DYSHIESGVLSGILK..TAQALVANLLTGFES..HFL
MS2   51 ELLTKHPTLGSGNDEATRRTLAIAKLREANDRCGQINREGFLHDKSLSWDPDVLQT....SIRSLIGNLLSGYRS..SLF
PP7   54 EPFSRKEEVPDSVXRASLIAFA.....NAEHNCRVMNQCGRY..YSVVNE.ELELLFSEASRLAKILISRWLSD.FWPDW

SP   148 LRCRFSGGATTSVNRLHGHPSWKHA.....CPQDVTKRAFKYLQAFKRACGDVVDLRVN.........EVRTSNKAVT VP
Qß   150 RHCRFSGGATTTNNRSYGHPSFKFA.....LPQACTPRALKYVLALR..ASTHFDTRIS.........DISPFNKAVT VP
GA   125 NDCSFSNGASQGFKLRDAAPFKKIA.....GQATVTAPAYDIAVAAVKTCAPWYAYMQETYGDETKWFRRVYGNGLFS VP
MS2  125 GQCTFSNGASMGHKLQDAAPYKKFA.....EQATVTPRALRAALLVRDQCAPW..IRHAVRYNESYEFRLVVGNGVFT VP
PP7  125 EEARFTGGASRLSSRKFSLPALKLAGFSERGQLSITAPALPYYRLYREGITA...........SDRGYTIVDDSRFDF VA

SP   214 KNSKTDRCIAIEPGWNMFFQLGVGAVLRDRLRLWKIDLNDQSTNQRLARDGSLLNHLATIDLSAASDSISLKLVELLMPP
Qß   214 KNSKTDRCIAIEPGWNMFFQLGIGGILRDRLRCWGIDLNDQTINQRRAHEGSVTNNLATVDLSAASDSISLALCELLLPP
GA   200 KNNKIDRAACKEPDMNMYLQKGAGSFIRKRLRSVGIDLNDQTRNQELARLGSIDGSLATIDLSSASDSISDRLVWDLLPP
MS2  198 KNNKIDRAACKEPDMNMYLQKGVGAFIRRRLKSVGIDLNDQSINQRLAQQGSVDGSLATIDLSSASDSISDRLVWSFLPP
PP7  194 KTAKAVRFIAMEPELNMLLQKSVGDTIRAALRKAGIDLNTQRLNQDLAYHGSVFRNLGTIDLSSASDTLSIELVRQYLPK

SP   294 EWYDLLTDLRSDEGILPDGRVVTYEKISSMGNGYTFELESLIFAAIARSVCELLEIDQSTVSVYGDDIIIDTRAAAPLMD
Qß   294 GWFEVLMDLRSPKGRLPDGSVVTYEKISSMGNGYTFELESLIFASLARSVCEILDLDSSEVTVYGDDIILPSCAVPALRE
GA   280 HVYSYLARIRTSFTMI.DGRLHKWGLFSTMGNGFTFELESMIFWALSKSIMLSMGVTGS.LGIYGDDIIVPVECRPTLLK
MS2  278 ELYSYLDRIRSHYGIV.DGETIRWELFSTMGNGFTFELESMIFWAIVKATQIHFGNAGT.IGIYGDDIICPSEIAPRVLE
PP7  274 RFLRYVLDLRTPYTSV.GGKKHRLEKVASMGNGFIFELQSLIYAAFAHAMTLVVGGRECDIAIYGDDIIVSECVVEPLMQ

SP   374 VFEYVGFTPNRKKTF.CDGPFRESCGKHWFQGVDVTPFYIRRPIRCLADMILVLNSIYRWGTVDGIWDP RALTVYEKYLK
Qß   374 VFKYVGFTTNTKKTF.SEGPFRESCGKHYYSGVDVTPFYIRHRIVSPADLILVLNNLYRWATIDGVWDP RAHSVYLKYRK
GA   358 VLSAVNFLPNEEKTF.TTGYFRESCGAHFFKDADMKPFYCKRPMETLPDVMLLCNRIRGWQTVGGMSDP RLFPIWKEFAD
MS2  356 ALAYYGFKPNLRKTF.VSGLFRESCGAHFYRGVDVKPFYIKKPVDNLFALMLILNRLRGWGVVGGMSDP RLYKVWVRLSS
PP7  353 FLEWHGFCPNLDKSYWGGDPFRESCGKHYFAGRDVTPVYVKGALDNLPALFRLFNSLKRWEEQTGIRIP DTIALVLSYIP

SP   453 LLPRNWRRNRIPDGYGDGALVGLATTNPFVIVKNYSRLYPVLVEVQRDVKRSEEGSYLYALLRDRETRYSPFLRDADRTG
Qß   453 LLPKQLQRNTIPDGYGDGALVGSVLINPFAKNRGWIRYVPVITDHTRDRERAELGSYLYDLFSRCLSESNDGLPLRGPSG
GA   437 MIPPKFKGGCNLD..RDTYLVSPDKP.....................GVSLVR..............IAKVRSGFNHAFPYG
MS2  435 QVPSMFFGGTDLA..ADYYVVSPPTAVSVYTKTPYGRLLADTRTSGFRLAR..............IARERKFFSEK....
PP7  433 KRDRVL....VPKTYSITAGLHFPAKGCVFPRTIYVRRYQRLIRKGRYMTEEQVD......ISKRLDDEVRYVDWLRNPP

SP   533 FDEAPL...........ATSLRRKTGRYKVAWIQDSAFIRPPYLITGIPEVKLAS.... 576
Qß   533 CDSADLFAIDQLICRSNPTKISRSTGKFDIQYIACSSRVLAPY..GVFQGTKVASLHEA 589
GA   482 HENGRYVHWLHMGSGEVLETISSARYRC....KPNSEWRTQIPLFPQELEACVLS.... 532
MS2  495 HDSGRYIAWFHTG.GEITDSMKSAGVRV....IRTSEWLTPVPTFPQECGPASSPR... 545
PP7  503 EALLPLEVWRRFSHRARHHGLPRERKRYRCVSSLEASSVWSSYEDWGEMA..... 552
```

FIG. 4. *Amino acid alignment of the replicases of PP7 and one representative of each of the four coliphage groups. The region with high homology is shown in brackets. Conserved amino acids are shown in bold face.*

determined by structure probing, comparative analysis, and translation predictability of mutant genotypes (Skripkin *et al.*, 1990; de Smit and van Duin, 1990). In Fig. 6 we compare the secondary structure of PP7 for this region with that of Q_β and with fr, a group I phage with high homology to MS2 (Adhin *et al.*, 1990a). Of all known group A phages, the fr structure shows most similarity with PP7, in particular the middle stem–loop which contains the coat gene start codon. Both PP7 and fr have a large internal loop, where MS2 only has a mismatch.

The maturation protein gene is terminated by a small hairpin that is conserved in all coliphages (Fig. 6). It contains the stop codon for the maturation protein, but its function is not known. As shown in Fig. 6, this stem–loop structure is conserved in PP7, but it does not contain the stop codon, which is located further downstream, in overlap with the coat start. We have noted that the PP7

TABLE 2

PERCENTAGE AMINO ACID IDENTITY BETWEEN THE REPLICASE PROTEINS IN THE CONSERVED REGION INDICATED IN FIGURE 4

	MS2	GA	Q_β	SP
PP7	45 (50)	42 (49)	42 (51)	43 (51)
MS2	—	68 (64)	49 (51)	52 (54)
GA	—	—	46 (50)	49 (52)
Q_β	—	—	—	73 (70)

Note. The identity at the nucleotide level is indicated in parentheses.

hairpin carries two nonsense triplets, but repeated sequencing in the 3′ terminal region of the maturation gene did not provide any evidence that any of these codons were in frame.

The coat initiator hairpin of MS2 RNA has been the subject of extensive research, which has revealed that ribosomes do not initiate by recognizing this helix. Rather, they bind to this RNA sequence when unfolded by thermal motion. It was found that the $\Delta G°$ of the hairpin is set at a value that just allows maximal ribosome loading (de Smit and van Duin, 1994). Evolutionary studies on disrupted initiator hairpins in phage MS2 have indeed shown that selection pressure operates neither on the exact shape of the hairpin nor on the sequence, but solely on the $\Delta G°$ of the structure (Olsthoorn *et al.*, 1994). We suppose that the PP7 stem–loop structure serves a similar purpose.

The second coat gene hairpin is also conserved in PP7. On its 3′ side we find in both phages a small region of sequence identity (underlined) that can serve as reference point to align the structures.

We note that small hairpins are found preferentially at the beginning of the coat and replicase genes and their presence seems a way to combine secondary structure

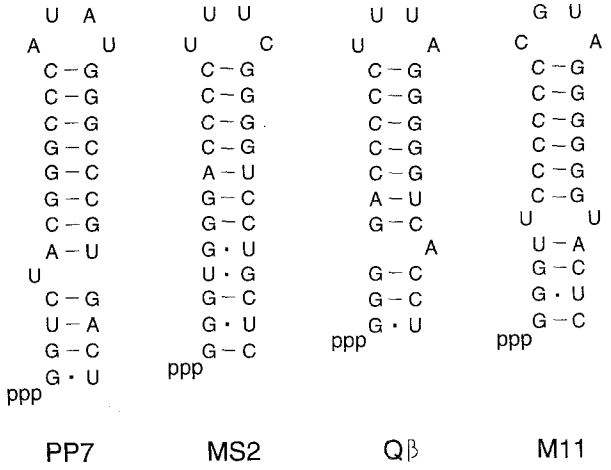

FIG. 5. The 5′ terminal hairpins of PP7, MS2, Q_β, and M11. M11 is a group III coliphage that was recently sequenced (Beekwilder, Nieuwenhuizen, and van Duin, unpublished observations).

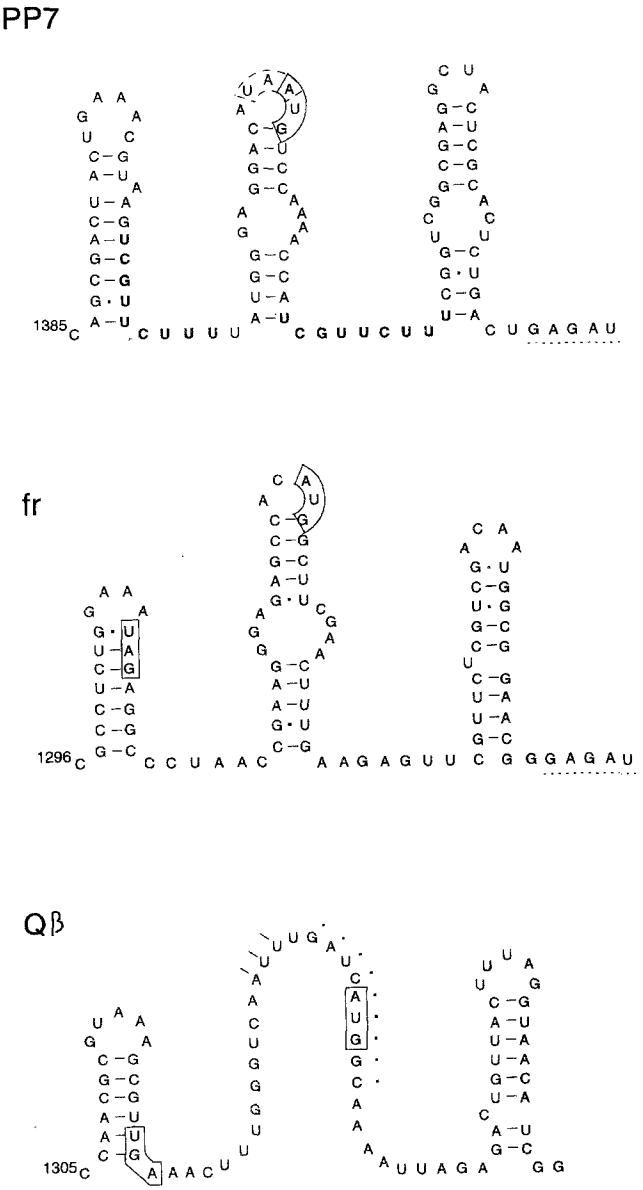

FIG. 6. Secondary RNA structure at the start of the coat protein genes in PP7, group I coliphage fr, and Q_β. Helices drawn for fr and Q_β are supported by phylogenetic comparison and structure probing (Skripkin *et al.*, 1990; Skripkin and Jacobson, 1993). (In fact, not fr but the closely related RNA of MS2 was probed). The AUG start codon and the stop codon of the maturation gene are indicated by a solid box. In PP7 the stop codon of the maturation gene overlaps the coat protein start and is indicated by a dashed box. A short stretch of sequence homology between PP7 and fr is indicated by a dotted line. A nine nucleotide repeat in PP7 is indicated in bold lettering. In the Q_β RNA structure possible loop sequences for the initiator hairpin are marked by slashes or dots.

with accessibility to ribosomes. Hairpins in the coding regions are generally much longer (Skripkin *et al.*, 1990; Skripkin and Jacobson, 1993).

For comparison we show the corresponding region for Q_β. The terminator and second coat gene hairpin that

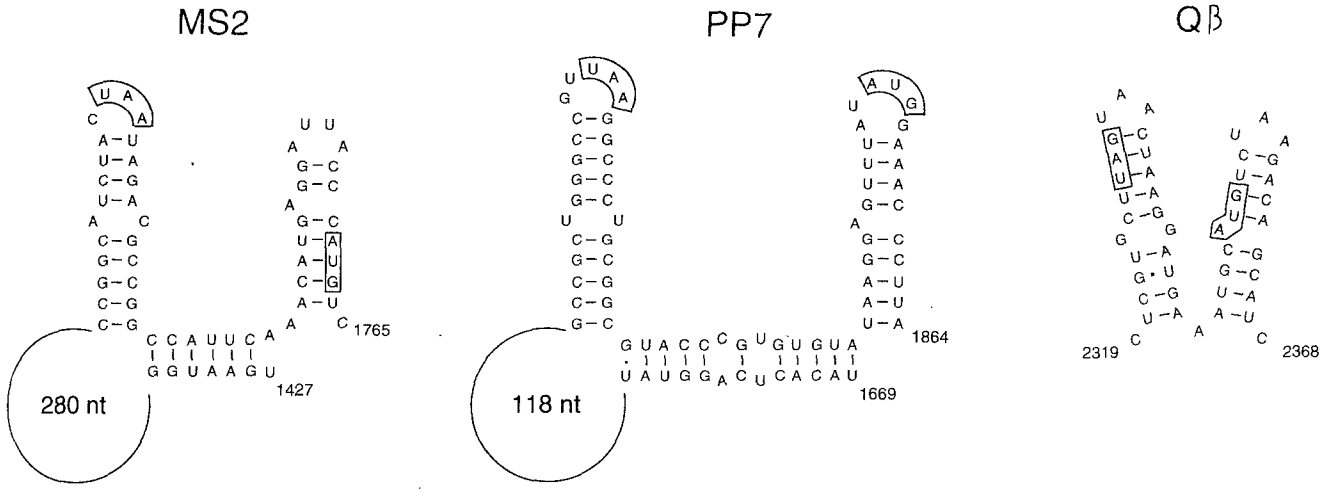

FIG. 7. RNA structure around the replicase start in PP7, MS2, and Q_β. Start codons and the stop codon of the preceding gene are indicated by a solid box. Structures shown for MS2 and Q_β are supported by phylogenetic comparison and structure probing (Skripkin *et al.*, 1990; Skripkin and Jacobson, 1993).

are shown are confirmed by structure probing and by phylogenetic comparison and they strongly resemble the structures present in PP7 and fr. Due to lack of sequence variation there is no certainty about the position of the initiator stem–loop in Q_β. Two possible loop positions are indicated by slashes or dots (see also Skripkin and Jacobson, 1993).

Replicase start. In the group A phages the replicase gene is preceded by the coat gene, while in the group B phages it is adjacent to the readthrough gene (Fig. 1). For this reason some differences in RNA structure between groups A and B exist at this point.

In Fig. 7 we compare the secondary structure of MS2, PP7, and Q_β at the start of the replicase cistron. The coat (readthrough) terminator hairpins are remarkably similar in all phages, in particular the conspicuous presence of the mismatch in the middle of the stem and the stop codon in or close to the loop. MS2 and PP7 have the contiguous G·C pairs in the lower part of the stem in common. Another striking similarity between these two phages is the nucleotide stretch between terminator and replicase initiator stem. In group A this stretch pairs with an internal segment of the coat coding region to form a long-distance interaction that is responsible for the translational coupling between coat and replicase genes in MS2 (Min Jou *et al.*, 1972; van Himbergen *et al.*, 1993). For PP7 we tentatively propose a similar interaction here. Obviously, in Q_β an equivalent interaction at this region is not possible, but Skripkin and Jacobson (1993) have proposed other regions of complementarity between the coat coding sequence and the vicinity of the replicase start to explain the translational polarity that was also shown to exist in this phage for these two genes.

The most interesting structure, however, in this region is the hairpin that contains the start codon of the replicase gene. In the coliphages this hairpin fulfills a key regulatory function. It can bind a coat protein dimer with high affinity, thereby preventing further replicase synthesis. This RNA–protein complex is also a first step in capsid formation (Witherell *et al.*, 1991).

In the coliphages the position of the bulged A, the identity of the loop nucleotides, and the number of base pairs in the stem are important for recognition by the coat protein (Witherell *et al.*, 1991). MS2 and Q_β have solved this repression problem in different ways. In MS2 the start codon is placed on the right side of the stem, whereas Q_β has it at the left side. In addition, the distance of the bulge from the loop and the sequences in the loop are different in both phages. PP7 provides another solution; the replicase start is positioned at the top in a hexanucleotide loop. The bulged A is placed as in the coliphages on the 5′ side of the stem and as in Q_β it occurs after the fourth pair from the top. On the other hand the identity of the bulge nucleotide is the same in PP7 and MS2, i.e., it is the A of the Shine-Dalgarno region.

A final point of interest concerns the differential stabilities of the terminator and initiator hairpins, specifically in MS2 and PP7. In PP7 these values are −19 and −6 kcal/mol, respectively. The much lower stability of the start helix is consistent with the finding that ribosomes are generally less able to initiate at structures with a stability stronger than approximately −7 kcal/mol (de Smit, 1994; Kubo *et al.*, 1990). Note that the structure depicted for MS2 (−3 kcal/mol) is weak enough to allow unlimited initiation in the absence of the coat protein and the long-distance interaction (Berkhout, 1986).

The 3′ untranslated region. The length of the 3′ untranslated region (UTR) varies in different phages: In group I it contains some 175 nts, whereas in the other three groups this number is close to 100. PP7 has the

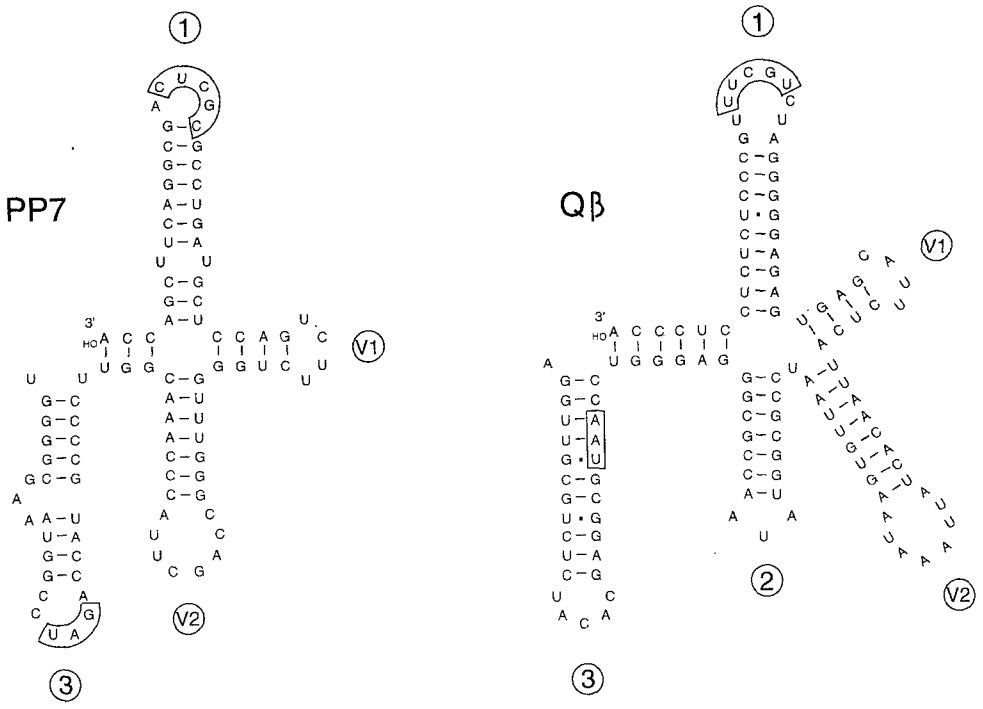

FIG. 8. Structure of the 3′ untranslated region of PP7 and Q_β. The replicase stop codon and the conserved sequence in the loop of helix 1 are boxed. The Q_β structure is supported by comparative analysis with about 10 other sequences (Adhin *et al.*, 1990b; Beekwilder, Nieuwenhuizen, and van Duin, submitted for publication).

shortest 3′ UTR with 81 nucleotides. The extra length of MS2 and fr is accounted for by the presence of three to four extra stem–loop structures as compared to the other groups (Adhin *et al.*, 1990b). The 3′ UTR of the coliphages folds into a separate domain bounded by a long-distance interaction between the last six nucleotides of the RNA and its complement present just downstream of the replicase terminator hairpin (Fig. 8). PP7 has a 3′ domain structure most resembling that of group III phages, from which helix 2 is absent. Specific similarities between PP7 and Q_β are stem–loop V1 with four base pairs and a tetraloop. These two features are also found in five other group III phages for which the sequence in this part is known. Hairpin V2 is a general feature of all coliphages and has a characteristic large loop usually rich in A residues. This hairpin contains the binding sites for ribosomal protein S1 and host factor (HFI) as well as for the maturation protein (Goelz and Steitz, 1977; Senear and Steitz, 1976; Shiba and Suzuki, 1981; Adhin *et al.*, 1990b). Hairpin 1 is present in all coliphages and it may have significance that the conserved loop sequence UGCUU is also present in PP7 albeit in the slightly changed form CGCUC.

The domain closing interaction in PP7 is only three vs six pairs in the other phages. We also note that PP7 has only two C residues at its 3′ end. This property is matched by the double G's at the 5′ end (Fig. 2). Because of the high resemblance with the Q_β structure we think our proposal for the PP7 3′ domain is realistic.

The coding regions. Of all proteins encoded in the RNA phages, the replicases show the largest similarity and nucleotide alignments between replicase genes are rather successful in regions where protein homology exists. In principle such homology would allow one to examine if RNA secondary structure is also conserved in the coding regions. Though this is an extensive project, that falls outside the scope of this paper, we provide two examples of a local analysis that will probably turn out to be representative for large parts of the complete molecule. One example shows how the RNA is folded around the motif YGDD, universally conserved in all RNA-dependent RNA replicases examined thus far. The nucleotide sequences around this motif can be aligned in all RNA bacteriophages. Comparative analysis supported by structure probing has enabled us to show the conservation of the folding of this region in A and B type coliphages and we can now compare these structures with PP7 (Fig. 9). Surprisingly, all coliphages form a hairpin with this sequence that shows overall similarity, but is different in detail. It is remarkable that even in PP7 a similar hairpin can form at virtually the same position in the RNA. However, among all these phages the sequence is used in a progressively shifted way. In Q_β the second aspartic acid codon is 4 nt away from the center of the loop, in MS2 this distance is two, in MX1 one nucleotide, while finally in PP7 the distance is zero. One can also visualize the gradual changes affecting this helix by focusing on the first isoleucine codon. This triplet

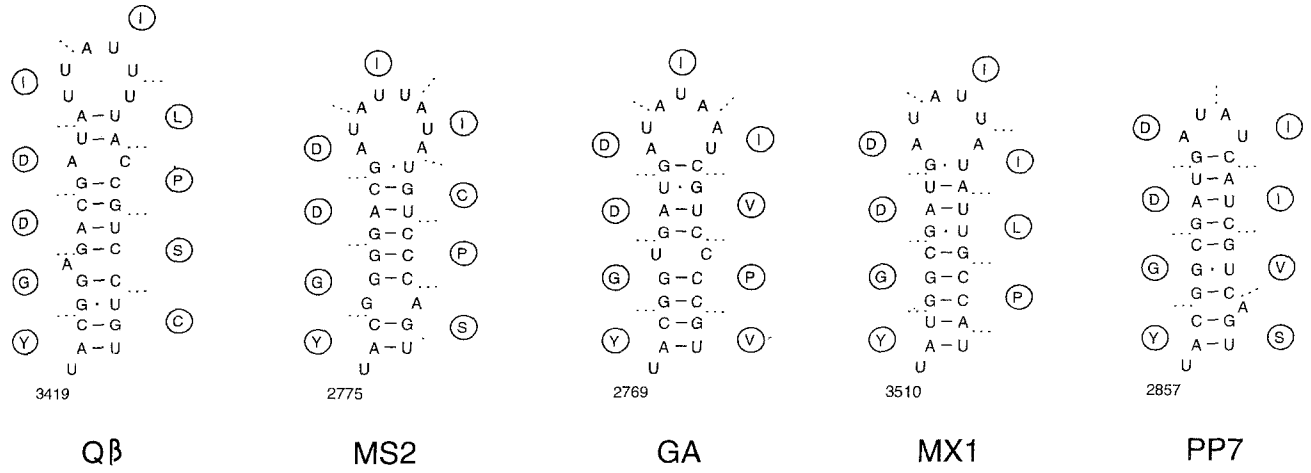

FIG. 9. RNA secondary structure at the amino acid motif YGDD which is conserved in all RNA-dependent RNA replicases. Structures are shown for Q_β, MS2, GA, MX1, and PP7. MX1 is a group III coliphage that was recently sequenced (Beekwilder, Nieuwenhuizen, and van Duin, unpublished observations). Dashes indicate the codons and the corresponding one letter amino acid code is circled.

lies on the left side of the helix in Q_β but going to the other phages one sees a progressive shift in position until this codon ends up at the 3′ side of the stem–loop structure in PP7. This flexibility in how to use different parts of aligned sequences to build a common RNA structure is a general feature in the RNA phages and frequently limits the information one can extract from alignments by conventional computer programs.

The second example is more extreme and concerns a hairpin centered around position 2812 in Q_β RNA, which is part of the internally located replicase binding site, the M2b site (Meyer *et al.*, 1981). The amino acids in this hairpin are centered around PP7 number 130 in the alignment of Fig. 4 and homology between all proteins is evident. The corresponding stem–loop structures for Q_β, MS2, and PP7 are shown in Fig. 10. Both the MS2 and Q_β structures are supported by structure probing and phylogenetic comparison (Skripkin *et al.*, 1990; Skripkin and Jacobson, 1993). In MS2 the hairpin is formed from a different part of the aligned sequences than in Q_β, but PP7 conforms rather well to the Q_β structure with the common GGA protein motif on the 3′ side of the helix. In a more extreme way, this example shows that, although several hairpins in coding regions can be shown to be conserved, they do not necessarily occupy an identical position in the aligned sequences in the different phage groups. More work is needed to determine if such shifts are corrected elsewhere in the chain, so as to bring the different foldings back into register with the alignment.

Replicase binding sites. When Q_β replicase is incubated with Q_β RNA to form a preinitiation complex, two internal regions are protected against RNase. One of these, the S-site, overlaps the ribosome binding site of the coat protein gene. The other one, the M-region, consists of three noncontiguous fragments, M5, M2b, and M11 (Meyer *et al.*, 1981). The M2b region is about 180 nucleotides long and at least part of its secondary structure is conserved in all four coliphages (Skripkin *et al.*, 1990; Skripkin and Jacobson, 1993). We therefore examined its potential preservation in PP7. Its global position in the sequence could be inferred from the amino acid alignment of Fig. 4, where the M2b site is found around position 110.

In Fig. 11 we present the existing base pairing models for SP and MS2 for a characteristic substructure of the M2b region. These models are supported by comparative evidence and structure probing (Skripkin *et al.*, 1990; Skripkin and Jacobson, 1993; Beekwilder and van Duin, unpublished observations). We note that the West stem–loop structure of the MS2 cloverleaf has been replaced

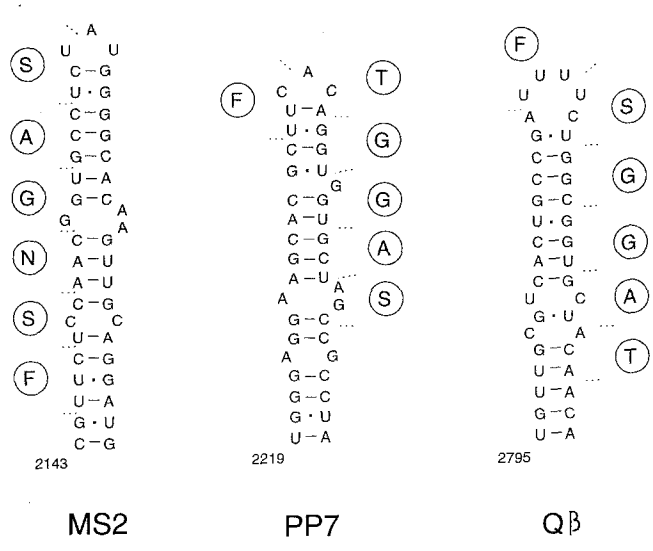

FIG. 10. RNA secondary structure at the conserved amino acid motif G(N)GAS(T); see alignment of Fig. 4 at amino acid numbers 166–170 (SP numbers).

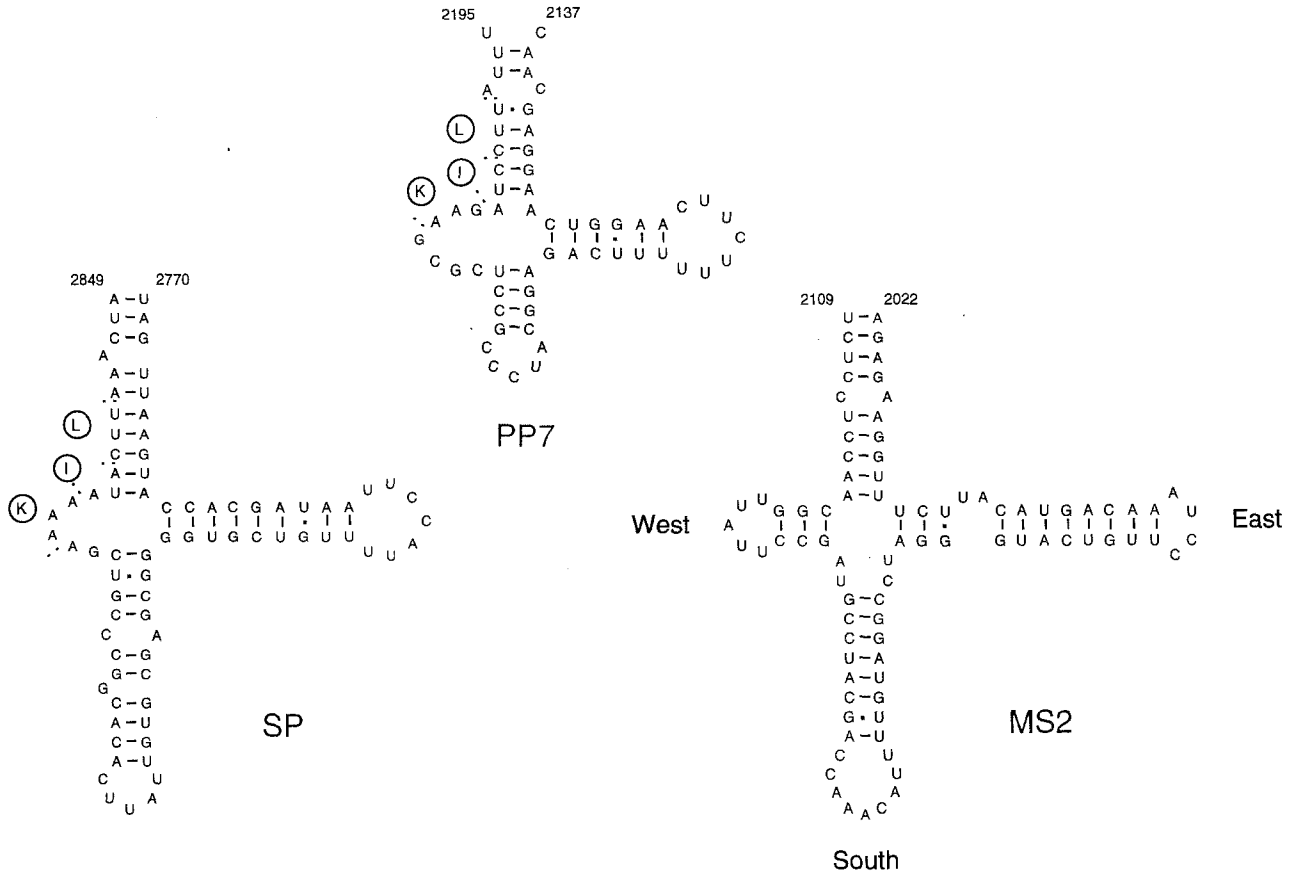

FIG. 11. RNA secondary structure at a substructure of the M2b replicase binding site. Amino acids that align quite well in the structures of SP and PP7 are circled. SP is chosen rather than Q_β because the protein homology between SP and PP7 is clearer for this part of the sequence. The three hairpins are referred to as East, South, and West, respectively.

by a hexanucleotide in SP. Another relevant feature is the large loop in the Southern hairpin and a short one in the East arm in MS2, while these characteristics are reversed in SP.

Our tentative proposal for PP7's equivalent of the M2b substructure shows good overall similarity with the coliphages, except that the East and South arms are substantially shorter. Further comparison reveals that PP7 shows more likeness to SP than to MS2. There is the replacement of the West arm by seven nucleotides and also the sizes of the loops in the two remaining helices conform those of SP, in particular the large pyrimidine-rich loop in the East helix. Similarly, there is an abundance of G and C residues in the Southern stem, in particular at the 3′ side.

The regular computer program FOLD only predicts the South arm of PP7, but when the energy parameters are slightly changed the East arm also appears. We also feel that the structure as we present it here is supported by the equivalent position of the string of aligned nucleotides encoding the amino acids K, I, and L in both SP and PP7 (Fig. 11).

The helix that we showed above in Fig. 10 is also part of the M2b site and it is quite suggestive that also in that example PP7 folds like group B rather than like group A.

Codon usage and base composition

In the coliphages codon choice in the coat gene closely follows that used by the host for highly expressed genes, while this preference tends to be random in the maturation and replicase genes.

With the aid of the codon preference table of West and Iglewski (1988) we have compared codon usage in the three major genes of PP7 with that of its host. With a few exceptions codon choice in the coat gene corresponds well to that of the host, while this preference is almost absent for the maturation and replicase genes. For instance, notoriously rare codons in Pseudomonas are absent from the coat gene, which is expected for a highly expressed gene, but these codons are not avoided in the maturation and replicase cistrons. An exception to the rule that the coat gene follows the codon bias of the host is the high frequency with which codons ending in U are used at the expense of those ending in C. For

example, in the PP7 coat gene, 31% of all ACN codons are ACU, while this number is only 3% in Pseudomonas genes; for the CUN vs CUU codon these numbers are 27 and 2%, respectively. Presumably, these U ending codons are avoided in the host to satisfy the GC drift of the Pseudomonas genome, which contains 67% G+C (Palleroni, 1984). Since synonymous codons ending in U or C are generally read by the same tRNA, an adverse effect on translation by using XYU instead of XYC is not likely.

In this context we note that PP7, containing 54% of G+C, has not adopted the base composition of the host. The evolutionary disadvantage here may be too strong a secondary structure. Consistent with this idea is that the G+C content of PP7 is only slightly higher than that of the coliphages which varies between 46 and 52% G+C.

CONCLUSIONS

Relatedness of PP7 to RNA coliphages

We have determined the sequence of the single-stranded RNA phage PP7 which grows in *P. aeruginosa* and have used this information for a molecular comparison with the RNA coliphages. The analysis shows that PP7 resembles the group I coliphages like MS2 in genome length and in genetic organization. Judged by sequence homology, however, PP7 is as distantly related to MS2 as it is to Q_β (Table 2). Despite the genetic closeness of PP7 to MS2 some features are more similar to Q_β type phages. For instance, the 3′ untranslated region of PP7 is much shorter than that of MS2 and close to the size of Q_β. Interestingly, also the folding of this region resembles that of Q_β and not that of the group A phages. Another point of similarity to Q_β is the structure of the hairpin at the start of the replicase cistron, which constitutes a high affinity coat protein binding site. In both PP7 and Q_β the bulged A nucleotide, believed important for protein recognition, is placed 4 bp from the loop, while this distance is only two in MS2. These similarities to both major coliphage groups suggests that PP7 and the coliphages branched off from a common ancestor before groups A and B separated. The possibility that PP7 is the result of horizontal transmission of group I coliphages to Pseudomonas leaves the Q_β traits of PP7 unexplained.

An archaetype structure for the RNA bacteriophages?

Except for the central portion of the replicase gene, nucleic acid sequence homology between PP7, Q_β, and MS2 is close to random. Nevertheless, our search has suggested that a substantial number of RNA structure elements are conserved. These include strategic regions such as translational start sites, the 5′ and 3′ untranslated sequences, as well as a replicase binding area. However, also in regular coding regions, only two examples of which have been presented in this paper (Figs. 9 and 10), structure homology can be identified. The diversity in genotypes seems of little concern to the phages as long as it leads to a conserved phenotype, i.e., secondary structure. In time it may become clear that, like other RNA species such as tRNA and ribosomal RNA, also the RNA phages possess a universally conserved ground structure.

ACKNOWLEDGMENTS

We thank Jules Beekwilder for communicating partial M11 and MX1 sequences. The Utah component thanks Ray Gesteland for advice and support. G.G. thanks Patrick Dhaese and Marc van Montagu for lab hospitality at the initiation of this project. Part of this work was supported by a grant from the Irish National Board for Science and Technology to J.F.A., by the Howard Hughes Medical Institute, by Grant GM48152 from the U.S. National Institutes of Health.

REFERENCES

ADHIN, M. R., and VAN DUIN, J. (1990). Scanning model for translation reinitiation in eubacteria. *J. Mol. Biol.* **213,** 811–818.

ADHIN, M. R., AVOTS, A., BERZIN, V., OVERBEEK, G. P., and VAN DUIN, J. (1990a). Complete nucleotide sequence of the group I RNA bacteriophage fr. *Biochim. Biophys. Acta* **1050,** 104–109.

ADHIN, M. R., ALBLAS, J., and VAN DUIN, J. (1990b). Secondary structure at the 3′ terminal region of RNA coliphages: Comparison with tRNA. *Biochim. Biophys. Acta* **1050,** 110–118.

ATKINS, J. F., and GESTELAND, R. F. (1983). Resolution of the discrepancy between a gene translation termination codon and the deduced sequence for release of the encoded polypeptide. *Eur. J. Biochem.* **137,** 509–516.

ATKINS, J. F., GESTELAND, R. F., REID, B. R., and ANDERSON, C. W. (1979a). Normal tRNAs promote ribosomal frameshifting. *Cell* **18,** 1119–1131.

ATKINS, J. F., STEITZ, J. A., ANDERSON, C. W., and MODEL, P. (1979b). Binding of mammalian ribosomes to MS2 phage RNA reveals an overlapping gene encoding a lysis function. *Cell* **18,** 247–256.

AXELROD, V. D., BROWN, E., PRIANO, C., and MILLS, D. R. (1991). Coliphage Q_β replication: RNA catalytic for single-strand release. *Virology* **184,** 595–608.

BARRERA, I., SCHUPPLI, D., SOGO, J. M., and WEBER, H. (1993). Different mechanisms of recognition of bacteriophage Q_β plus and minus strand RNAs by Q_β replicase. *J. Mol. Biol.* **232,** 512–521.

BASSEL, B. A., and MILLS, D. R. (1979). Initiation of translation with *Pseudomonas aeruginosa* phage PP7: nucleotide sequence of the coat cistron ribosome binding site. *Nucleic Acis Res.* **6,** 2003–2016.

BEREMAND, M. W., and BLUMENTHAL, T. (1979). Overlapping genes in RNA phage: A new protein implicated in cell lysis. *Cell* **18,** 257–266.

BERKHOUT, B. (1986). "Translational Control Mechanisms in RNA Bacteriophage MS2," Ph.D. thesis. University of Leiden.

BERKHOUT, B., DE SMIT, M. H., SPANJAARD, R. A., BLOM, T., and VAN DUIN, J. (1985). The amino terminal half of the MS2-coded lysis protein is dispensable for function: Implications for our understanding of coding region overlaps. *EMBO J.* **4,** 3315–3320.

BERKHOUT, B., SCHMIDT, B. F., VAN STRIEN, A., VAN BOOM, J. VAN WESTRENEN, J., and VAN DUIN, J. (1987). Lysis gene of bacteriophage MS2 is activated by translation termination at the overlapping coat gene. *J. Mol. Biol.* **195,** 517–524.

BIEBRICHER, C. K., and LUCE, R. (1993). Sequence analysis of RNA species synthesized by Q_β replicase without template. *Biochemistry* **32,** 4848–4854.

BRADLEY, D. E. (1966). The structure and infective process of a *Pseudomonas aeruginosa* bacteriophage containing ribonucleic acid. *J. Gen. Microbiol.* **45,** 83–96.

BRADLEY, D. E. (1972). Shortening of *Pseudomonas aeruginosa* pili after RNA-phage adsorption. *J. Gen. Microbiol.* **72,** 303–319.

BRUENN, J. A. (1991). Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucleic Acids Res.* **19,** 217–226.

CHEN, E. Y., and SEEBURG, P. H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* **4,** 165–170.

COLEMAN, J., INOUYE, M., and ATKINS, J. F. (1983). Phage MS2 lysis protein does not require coat protein to mediate cell lysis. *J. Bacteriol.* **153,** 1098–1100.

CRAWFORD, E. M., and GESTELAND, R. F. (1964). The adsorption of bacteriophage R17. *Virology* **22,** 165–167.

DAVIES, J. W., and BENIKE, C. (1974). Translation of virus mRNA: Synthesis of bacteriophage PP7 proteins in cell-free extracts from *Pseudomonas aeruginosa*. *Virology* **61,** 450–457.

DAYHUFF, T. J., ATKINS, J. F., and GESTELAND, R. F. (1986). Characterization of ribosomal frameshift events by protein sequence analysis. *J. Biol. Chem.* **261,** 7491–7500.

DE SMIT, M. H. (1994). "Regulation of Translation by mRNA Structure," Ph.D. thesis. University of Leiden.

DE SMIT, M. H., and VAN DUIN, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. USA* **87,** 7668–7672.

DE SMIT, M. H., and VAN DUIN, J. (1994). Translational initiation on structured messengers. Another role for the Shine-Dalgarno interaction. *J. Mol. Biol.* **235,** 173–184.

DEVEREUX, I., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12,** 387–395.

DHAESE, P. (1979). "Studie van de RNA bacteriofagen PRR1 en PP7: Hun verwantschap met de RNA colifagen," Ph.D. thesis. University of Ghent.

DHAESE, P., VANDEKERCKHOVE, J. S., and VAN MONTAGU, M. (1979). The primary structure of the coat protein of the broad-host-range RNA bacteriophage PRR1. *Eur. J. Biochem.* **94,** 375–386.

DHAESE, P., LENAERTS, A., GIELEN, J., and VAN MONTAGU, M. (1980). Complete amino acid sequence of the coat protein of the *Pseudomonas aeruginosa* bacteriophage PP7. *Biochim. Biophys. Acta* **94,** 1394–1400.

FIERS, W., CONTRERAS, R., DUERINCK, F., HAEGEMAN, C., ISERENTANT, D., MERREGAERT, J., MIN JOU, W., MOLEMANS, F., RAEYMAKERS, A., VANDENBERGHE, A., VOLCKAERT, C., and ISEBAERT, M. (1976). Complete nucleotide sequence of bacteriophage MS2-RNA: Primary and secondary structure of the replicase gene. *Nature (London)* **260,** 500–507.

FURUSE, K. (1987). Distribution of coliphages in the environment: General considerations. *In* "Phage Ecology" (S. M. Goyal, Ed.), pp. 87–124. Wiley, New York.

GOELZ, S., and STEITZ, J. A. (1977). *E. coli* ribosomal S1 recognizes two sites in Q_β RNA. *J. Biol. Chem.* **252,** 5177–5184.

GOESSENS, W. H. F., DRIESSEN, A. J. M., WILSCHUT, J., and VAN DUIN, J. (1988). A synthetic peptide corresponding to the C-terminal 25 residues of phage MS2 coded lyses protein dissipates the protonmotive force in *Escherichia coli* membrane vesicles by generating hydrophilic pores. *EMBO J.* **7,** 867–873.

GOLMOHAMMADI, R., VALEGÅRD, K., FRIDBORG, K., and LILJAS, L. (1993). The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J. Mol. Biol.* **234,** 620–639.

HEUS, H. A., and PARDI, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253,** 191–194.

HOLLOWAY, B. W. (1969). Genetics of pseudomonas. *Bacteriol. Rev.* **33,** 419–433.

INOKUCHI, Y., TAKAHASHI, R., HIROSE, T., INAYAMA, S., JACOBSON, B., and HIRASHIMA, A. (1986). The complete nucleotide sequence of group II RNA coliphage GA. *J. Biochem.* **99,** 1169–1180.

INOKUCHI, Y., JACOBSON, A. B., HIROSE, T., INAYAMA, S., and HIRASHIMA, A. (1988). Analysis of the complete nucleotide sequence of the group IV RNA-coliphage SP. *Nucleic Acids Res.* **16,** 6205–6221.

ISERENTANT, D., VAN MONTAGU, M., and FIERS, W. (1980). Studies on the bacteriophage MS2, XLI. Nature of the azure mutation. *J. Mol. Biol.* **139,** 243–263.

KAJITANI, M., KATO, A., WADA, A., INOKUCHI, Y., and ISHIHAMA, A. (1994). Regulation of the *Escherichi coli hfq* gene encoding the host factor for phage Q_β. *J. Bacteriol.* **176,** 531–534.

KAMER, G., and ARGOS, P. (1984). Primary structural comparison of RNA dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* **12,** 7269–7282.

KARNIK, S., and BILLETER, M. (1983). The lysis function of RNA bacteriophage Q_β is mediated by the maturation (A2) protein. *EMBO J.* **2,** 1521–1526.

KUBO, M., HIGO, Y., and IMANOKA, T. (1990). Biological treshold values of prokaryotic gene expression which is controlled by the DNA inverted repeat sequence and the mRNA secondary structure. *J. Ferment. Bioeng.* **69,** 305–307.

MEKLER, P. (1981). "Determination of Nucleotide Sequences of the Bacteriophage Q_β Genome: Organization and Evolution of an RNA Virus," Ph.D. thesis. University of Zürich.

MEYER, F., WEBER, H., and WEISSMANN, C. (1981). Interactions of $Q\beta$ replicase with $Q\beta$ RNA. *J. Mol. Biol.* **153,** 631–660.

MIN JOU, W., HAEGEMAN, G., YSEBAERT, M., and FIERS, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature (London)* **237,** 82–88.

OLSTHOORN, R. C. L., LICIS, N., and VAN DUIN, J. (1994) Leeway and constraints in the forced evolution of a regulatory RNA helix. *EMBO J.* **13,** 2660–2668

POCH, O., SAUVAGET, I. DELARUE, M., and TORDO, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8,** 3867–3874.

PALLERONI, N. J. (1984). *In* "Bergey's Manual of Systematic Bacteriology" (N. R. Kreig and J. G. Holt, Eds.), Vol. 1, pp. 165–172. Williams and Wilkins, Baltimore.

PARANCHYCH, W., and FROST, L. A. (1988). The physiology and biochemistry of pili. *Arch. Microbiol. Physiol.* **29,** 53–114.

PEABODY, D. S. (1989). Translational repression by bacteriophage MS2 coat protein does not require cysteine residues. *Nucleic Acids Res.* **17,** 6017–6027.

SANGER, F., NICKLEN, S., and COULSON, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5476.

SENEAR, A. W., and STEITZ, J. A. (1976). Site specific interaction of Q_β host factor and ribosomal protein S1 with Q_β and R17 RNAs. *J. Biol. Chem.* **251,** 1902–1912.

SHIBA, R., and SUZUKI, Y. (1981). Localization of A protein in the RNA-A protein complex of RNA phage MS2. *Biochim. Biophys. Acta* **654,** 249–255.

SKRIPKIN, E. A., ADHIN, M. R., DE SMIT, M. H., and VAN DUIN, J. (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 JACOBSON, A. B. (1993). A two-dimensional model at the nucleotide level for the central hairpin of coliphage Q_β. *J. Mol. Biol.* **233,** 245–260.

TUERK, C., GAUSS, P., THERMES, C., GROEBE, D. R., GAYLE, M., GUILD, N., STORMO, G., D'AUBENTON-CARAFA, Y., UHLENBECK, O. C., TINOCO, I., BRODY, E., and GOLD, L. (1988). CUUCGG hairpins: Extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. USA* **85,** 1364–1368.

VALEGÅRD, K., LILJAS, L., FRIDBORG, K., and UNGE, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature (London)* **346,** 36–41.

VAN DUIN, J. (1988). Single stranded RNA bacteriophages. *In* "The Viruses" (H. Fraenkel-Conrat and R. Wagner, Eds.), pp. 117–167. Plenum, New York.

VAN HIMBERGEN, J., VAN GEFFEN, B., and VAN DUIN, J. (1993). Translational control by a long range RNA-RNA interaction; a basepair substitution analysis. *Nucleic Acids Res.* **21,** 1713–1717.

VOORMA, H. O., BENNE, R., and DEN HARTOG, T. J. A. (1971). Binding of aminoacyl-tRNA to ribosomes programmed with bacteriophage MS2. *Eur. J. Biochem.* **18,** 451–462.

WEST, S. E. H., and IGLEWSKI, B. H. (1988). Codon usage in *Pseudomonas aeruginosa. Nucleic Acids Res.* **16,** 9323–9335.

WINTER, R. B., and GOLD, L. (1983). Overproduction of bacteriophage Qβ maturation protein leads to cell lysis. *Cell* **33,** 877–885.

WITHERELL, G. W., GOTT, J. M., and UHLENBECK, O. C. (1991). Specific interaction between RNA phage coat proteins and RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **40,** 185–220.

YOUNG, R. (1993). Bacteriophage lysis: mechanism and regulation. *Microbiol. Rev.* **56,** 431–481.

ZINDER, N. E. (Ed.) (1975). "RNA Phages." Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

ZUKER, M., and STIEGLER, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9,** 133–148.