

Genome Structure of Caulobacter Phage phiCb5[∇]

Andris Kazaks, Tatyana Voronkova, Janis Rumnieks, Andris Dishlers, and Kaspars Tars*

Biomedical Research and Study Center, Ratsupites 1, LV 1067 Riga, Latvia

Received 28 October 2010/Accepted 3 February 2011

The complete genome sequence of caulobacter phage phiCb5 has been determined, and four open reading frames (ORFs) have been identified and characterized. As for related phages, the ORFs code for maturation, coat, replicase, and lysis proteins, but unlike other *Leviviridae* members, the lysis protein gene of phiCb5 entirely overlaps with the replicase in a different reading frame. The lysis protein of phiCb5 is about two times longer than that of the distantly related MS2 phage and presumably contains two transmembrane helices. Analysis of the proposed genome secondary structure revealed a stable 5' stem-loop, similar to other phages, and a substantially shorter 3' untranslated region (UTR) structure with only three stem-loops.

The small RNA phages belonging to the *Leviviridae* family have been extensively used as models to study various problems in molecular biology, including translational control, virus evolution, structure, and assembly. *Leviviridae* coliphages are divided into two genera, *Levivirus* and *Allolevivirus*. The levivirus genome (represented by phage MS2 in Fig. 1A) encodes four proteins: open reading frame 1 (ORF1) encodes a maturation or “A” protein (AP), responsible for the attachment of phages to bacterial F pili; ORF2 encodes the coat protein (CP); ORF3 encodes the replicase subunit (RP); and a fourth open reading frame partially overlaps with ORF2 and ORF3 and encodes a lysis protein (LP). Alloleviviruses (represented by phage Qβ in Fig. 1A) do not have a separate gene for the LP; instead, the AP is responsible for cell lysis. Capsids of alloleviviruses contain about 5% of A1 protein, which is a prolonged read-through variant of CP that has been shown to be necessary for infection (8). The genome organizations of the *Pseudomonas* phage PP7 (12) and broad-host-range phage PRR1 (14) are similar to that of leviviruses, while the *Acinetobacter* AP205 phage (10) has a slightly different genome organization (Fig. 1A).

The RNA phage φCb5, first isolated by Schmidt (16), infects bimorphic *Caulobacter crescentus* bacteria through adsorption to pili specific to swarmer cells (15). Phage φCb5 RNA was isolated and sequenced as previously described (13). The obtained sequence from several overlapping clones covered most of the phage genome, except the 5' and 3' ends. To determine the phage genome sequence of the 5' end, its cDNA was tailed with dATP using terminal transferase, and PCR was carried out using the 5'-GCGCG(T)₁₈ primer and a primer complementary to nucleotides (nt) 257 to 275. The PCR products were cloned, and eight clones were sequenced. In all cases, we obtained the same 5' sequence, with the exception of some shortened variants. To resolve the 3' end, a poly(A) tail was added to the phage RNA using poly(A) polymerase. The cDNA was synthesized with the 5'-GCGCG(T)₁₈ primer, and PCR was carried out using the 5'-GCGCG(T)₁₈ primer and a primer complementary to nucleotides 3142 to 3161. The PCR

fragment was cloned, and four clones were sequenced. All of the clones displayed the 3' end of the RP gene followed by 82 additional nucleotides.

The genome of φCb5 is organized in a way similar to that of leviviruses (Fig. 1A). After a short 5' untranslated region (UTR), ORF1 encodes AP, ORF2 encodes CP, and ORF3 encodes RP. However, the LP gene of φCb5 is placed differently and entirely overlaps with the RP gene in the (+1) reading frame.

The nucleotide sequence of the φCb5 genome and amino acid sequences of the individual proteins have very low homologies with their counterparts in other RNA phages. The only sequence which can be aligned unambiguously is the central part of RP (residues 295 to 537). The coat protein of φCb5 has low sequence similarity to other RNA phages, and none of the residues conserved among PRR1, PP7, and all coliphages are conserved in φCb5. However, reliable alignment based on known three-dimensional structures of coat proteins can be performed (13). In both cases, phylogenetic analysis suggests that φCb5 forms a distant branch among *Leviviridae* and does not belong to either *Levivirus* or *Allolevivirus* (Fig. 1B).

Like in other related phages, the capsid of φCb5 consists of 180 CP monomers. The crystal structure of the φCb5 capsid has been described in detail by Plevka et al. (13).

The CP of φCb5 is unusually short, only 122 amino acid residues, while the CPs of other *Leviviridae* phages have lengths ranging from 127 to 132 residues.

Although the most obvious initiation site for AP is the first AUG codon in the genome that also has a strong preceding Shine-Dalgarno (SD) sequence, mass spectrometry revealed the presence of a protein of a smaller mass than predicted from the sequence (data not shown). To establish the actual translation start site of the AP, the proteins of purified φCb5 virions were separated by SDS-PAGE, and the N-terminal sequence of the 40-kDa band was determined. The sequence was found to be ARIRN, corresponding to a nucleotide sequence 78 nucleotides from the 5' end of the genome. The sequence is immediately preceded by a UUG codon, which can serve as an initiation codon in bacteria, which is probably the case for the AP of φCb5. However, we cannot exclude the possibility that the upstream AUG codon is in fact used for translational initiation and that proteolytic cleavage occurs later.

* Corresponding author. Mailing address: Biomedical Research and Study Center, Ratsupites 1, LV 1067 Riga, Latvia. Phone: (371) 270-76237. Fax: (371) 674-42407. E-mail: kaspars@biomed.lu.lv.

[∇] Published ahead of print on 16 February 2011.

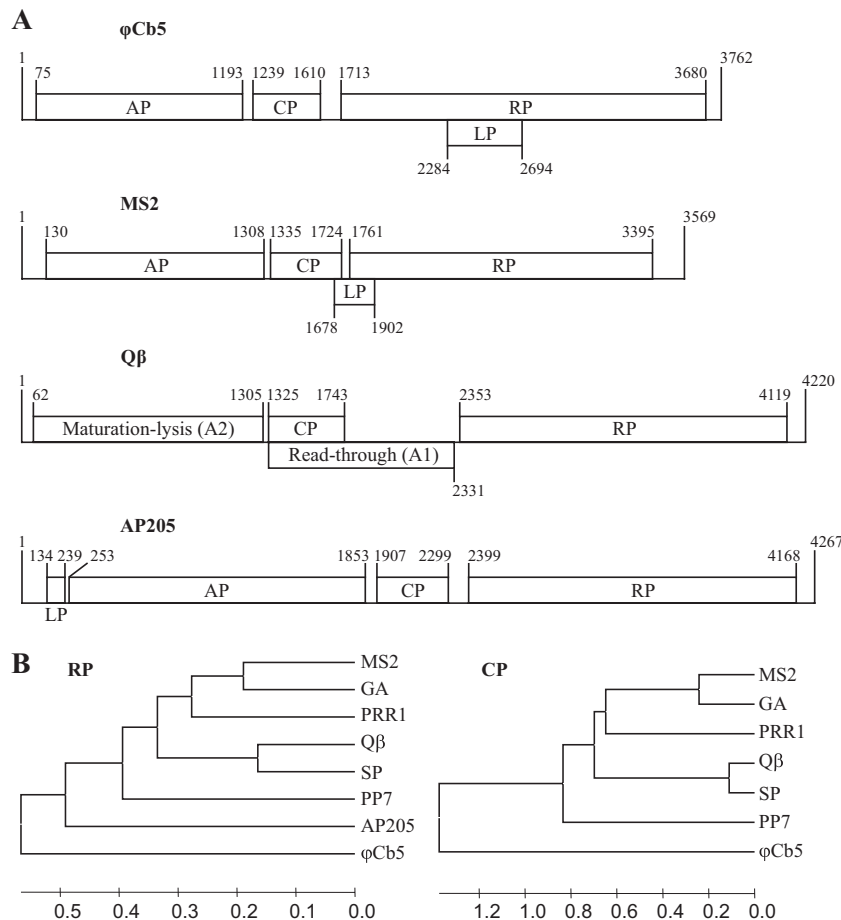


FIG. 1. (A) The genome organization of small RNA phages ϕ Cb5, MS2, Q β , and AP205. Genes are drawn to their approximate scale. (B) Unweighted-pair group method using average linkages (UPGMA) tree of small RNA phages, based on sequence alignment of the RP central part and structure-based alignment of CP. AP205 is not included in the CP alignment-based tree, since its structure is unknown and the CP sequence itself shows no significant similarity to other phages. Trees were constructed by MEGA (17).

In alloviruses, lysis is achieved by the maturation protein that blocks MurA, an enzyme in the pathway of murein biosynthesis (4). However, this is not the case for ϕ Cb5, since overexpression of the AP gene of ϕ Cb5 in *Escherichia coli* did not cause cell lysis (Fig. 2A). Leviviruses have dedicated lysis proteins which form pores in the cellular membrane, leading to activation of autolysins and eventually cell lysis (6). The sequences, lengths, and locations in the genome of the LPs vary among different leviviruses, and their only conserved features seem to be the clustering of positively charged residues near the N terminus and a hydrophobic region near the C terminus, which has been demonstrated to form a transmembrane helix in case of MS2 (6). In MS2, the last 30 residues of LP are necessary and sufficient for cell lysis (3), suggesting that the presence of positively charged residues in the N-terminal region of the protein is not crucial.

In ϕ Cb5, no obvious ORF corresponding to the LP of leviviruses or to AP205 could be detected. Analysis of the translated genome sequence using the TMHMM 2.0 server (11) revealed in total three transmembrane helices that entirely overlapped with the RP gene in a different reading frame. The first helix, encoded by nucleotides 2098 to 2226, lacked any suitable upstream initiation codon, and when the respective sequence was cloned into an

expression plasmid with an AUG initiation codon, no change in cell growth was observed upon induction (Fig. 2A). The other two transmembrane helices were found in a potential ORF that had a strong SD sequence but an unusual start codon, UUG. However, unusual LP start codons have been reported earlier for several other small RNA phages, like fr (1), PP7 (12), and PRR1 (14). Possibly, non-AUG start codons might help to limit expression of potentially dangerous proteins. The ORF was 135 residues long, considerably exceeding the LP lengths of other small RNA phages (Fig. 2B). Upon ORF cloning and expression, the optical density (OD) decreased (Fig. 2A) in a manner similar to that described for AP205 and PRR1 LPs (10, 14). The cell growth was not affected by a similar expression plasmid with an identical sequence except for a termination codon that was placed after the second codon of the putative LP gene (Fig. 2A). When both potential transmembrane helices were cloned and expressed separately, cell growth was halted, but no decrease in the OD value was observed (Fig. 2A). It should be noted that the above putative LP expression attempts were carried out in *E. coli*; therefore, we cannot exclude the possibility that analyzed ORFs would behave differently in ϕ Cb5 host *Caulobacter*.

RP is the most conserved protein among *Leviviridae* phages,

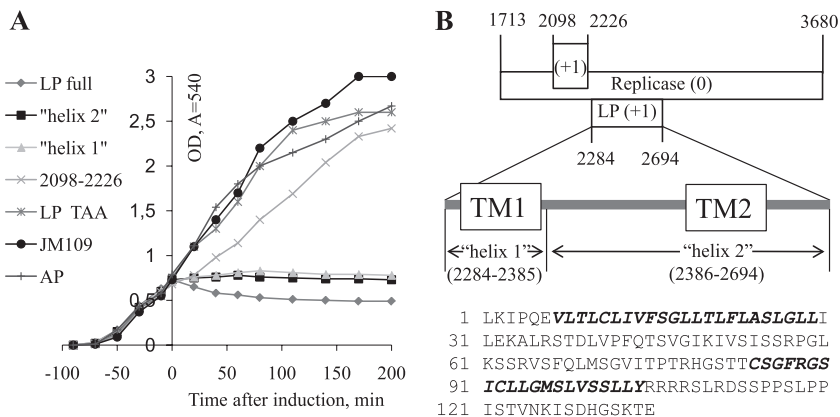


FIG. 2. Properties of LP. (A) *E. coli* cell growth upon expression of the proposed LP, the individual transmembrane helices 1 and 2 of LP, LP with early termination codon (LP TAA), AP, and the translated sequence of nucleotides 2098 to 2226 in the ϕ Cb5 genome. All the ORFs were cloned under arabinose-inducible promoter in pBAD plasmid (Invitrogen) and expressed in *E. coli* strain JM109. Growth of untransformed JM109 is shown as well. (B) Position of LP in the RP gene. Positions of predicted transmembrane regions are shown as "TM1" and "TM2" and actual expressed protein sequences as "helix 1" and "helix 2." The sequence of the proposed LP, with transmembrane regions in bold italic, is shown.

with the highest homology around the active site. Although the overall sequence identity is low, the ϕ Cb5 RP has the GDD and FRESCG motifs that are totally conserved among all known *Leviviridae* phages. There are several initiation codons in the beginning of the RP ORF. The first AUG also has a significant SD sequence upstream; therefore, we assume that RP is most likely translated from the first AUG codon.

The small RNA phages have a characteristic RNA stem-loop structure, including the RP start codon (Fig. 3A). This stem-loop serves as a binding site for a CP dimer, which represses the translation of the RP gene in late stages of infection. The putative secondary RNA structure of the region surrounding the RP initiation codon is shown in Fig. 3B. There are two more stem-loops close to the three downstream AUG codons that could potentially serve as repression sites (Fig. 3B). However, we failed to demonstrate binding of any of the three corresponding stem-loop RNA oligonucleotides (5'-UC AUCCCUAGCUUU AUGAGGCUAAGAUGA, 5'-UAUC AGGCAGUUAUGAAAGACUACCUGAUG, 5'-AGGACG UUGAGCGUGACAUGUCACGCCUCCAACUCCU) to CP using a filter binding assay as described for phage R17 (5). As none of the conserved RNA-binding residues in related phages were identified in ϕ Cb5 CP (13), the interactions of ϕ Cb5 CP with RNA may be very different.

The small RNA phages have a characteristic stable stem-loop structure at their 5' ends, believed to be necessary for strand separation during replication (2, 18). A similar loop is found near the 5' end of the ϕ Cb5 genome (Fig. 4A).

3' UTRs of small RNA phages are folded in a separate domain composed of four to nine stem-loops (Fig. 4B). ϕ Cb5 appears to have the simplest arrangement known so far, with just three stem-loops, including the RP termination codon containing the R1 loop. In coliphages and AP205, there is a conserved UGCUU sequence 15 to 17 nt from the 3' end that in the case of phage Q β has been shown to regulate replication via a long-distance interaction (9). The last stem-loop of ϕ Cb5 RNA is somewhat similar to the U1 loops in related phages, and it contains a UGCUG sequence 16 nt from the 3' end. A

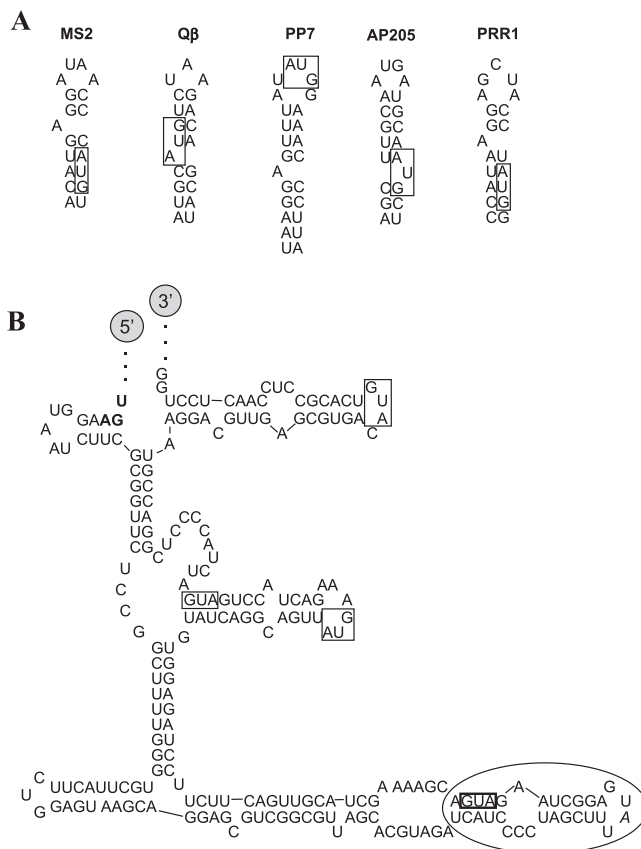


FIG. 3. (A) RP gene operator stem-loops of small RNA phages. RP initiation codons are boxed. Secondary structures were calculated using the RNAfold server (7). (B) Proposed secondary structure of the ϕ Cb5 genome region between the termination codon of the CP gene (shown in bold) and +124 in the RP gene. The first possible initiation codon of the RP gene is shown in a box with bold lines. The three downstream AUG codons in the same reading frame are boxed as well. The stem-loop structure containing the first initiation codon is circled.

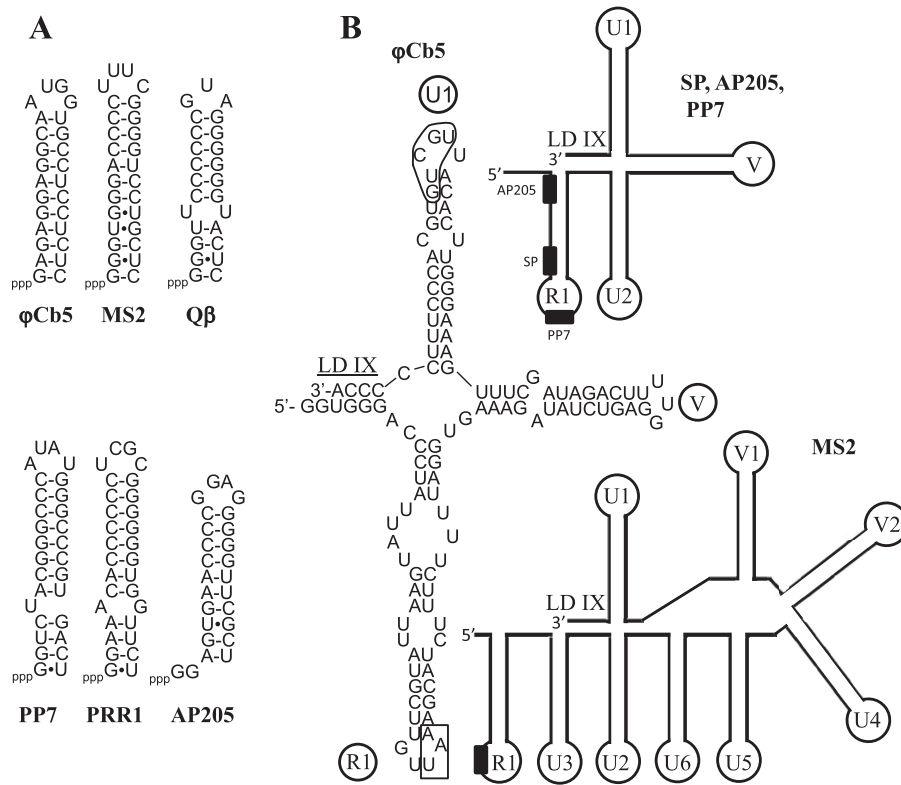


FIG. 4. Structure of terminal UTRs. (A) Hairpin loops near the 5' ends of genomes of small RNA phages. (B) Structure of 3' ends of small RNA phages. Sequence is displayed only for ϕ Cb5; for other phages, the schematic structures of stem-loops are shown. The RP termination codon in the R1 loop is boxed, and the conserved sequence in the U1 loop is circled. RP termination codons are shown as black boxes. Names of stem-loops and the long-distance interaction region (LD IX) are adopted from reference 10.

sequence complementary to UGCUG is found in two positions in the RP gene, but due to insignificant sequence similarity of ϕ Cb5 and Q β genomes, it cannot be concluded whether a long-distance interaction takes place in ϕ Cb5 as well.

Nucleotide sequence accession number. The sequence has been deposited in GenBank under accession number HM066936.

This study was supported by ESF grant 1DP/1.1.1.2.0/09/APIA/VIAA/150, ERDF grant 2DP/2.1.1.0/10/APIA/VIAA/052, and grant 09.1294 from the Latvian Council of Science.

We thank Pavel Plevka for participation in the RNA-binding measurements. Lars Liljas and Janis Klovinš are acknowledged for valuable discussions.

REFERENCES

1. Adhin, M. R., A. Avots, V. Berzin, G. P. Overbeek, and J. van Duin. 1990. Complete nucleotide sequence of the group I RNA bacteriophage fr. *Biochim. Biophys. Acta* **1050**:104–109.
2. Beekwilder, M. J., R. Nieuwenhuizen, and J. van Duin. 1995. Secondary structure model for the last two domains of single-stranded RNA phage Q β . *J. Mol. Biol.* **247**:903–917.
3. Berkhout, B., M. H. de Smit, R. A. Spanjaard, T. Blom, and J. van Duin. 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.
4. Bernhardt, T. G., I. Wang, D. K. Struck, and R. Young. 2001. A protein antibiotic in the phage Q β virion: diversity in lysis targets. *Science* **292**:2326–2329.
5. Carey, J., V. Cameron, P. L. de Haseth, and O. C. Uhlenbeck. 1983. Sequence-specific interaction of R17 coat protein with its RNA binding site. *Biochemistry* **22**:2601–2610.
6. Goessens, W. H. F., A. J. M. Driessen, J. Wilschut, and J. van Duin. 1988. A synthetic peptide corresponding to the C-terminal 25 residues of phage MS2

- coded lysis protein dissipates the protonmotive force in *Escherichia coli* membrane vesicles by generating hydrophilic pores. *EMBO J.* **7**:867–873.
7. Gruber, A. R., R. Lorenz, S. H. Bernhart, R. Neuböck, and I. L. Hofacker. 2008. The Vienna RNA Websuite. *Nucleic Acids Res.* **36**:W70–W74.
8. Hofstetter, H., H. Monstein, and C. Weissmann. 1974. The readthrough protein A1 is essential for the formation of viable Q β particles. *Biochim. Biophys. Acta* **374**:238–251.
9. Klovinš, J., and J. van Duin. 1999. A long-range pseudoknot in Q β RNA is essential for replication. *J. Mol. Biol.* **294**:875–884.
10. Klovinš, J., G. P. Overbeek, S. H. van den Worm, H. W. Ackermann, and J. van Duin. 2002. Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. *J. Gen. Virol.* **83**:1523–1533.
11. Krogh, A., B. Larsson, G. von Heijne, and E. L. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**:567–580.
12. Olsthoorn, R. C. L., G. Garde, T. Dayhuff, J. F. Atkins, and J. van Duin. 1995. Nucleotide sequences of a single-stranded RNA phage from *Pseudomonas aeruginosa*: kinship to coliphages and conservation of regulatory RNA structures. *Virology* **206**:611–625.
13. Plevka, P., et al. 2009. The structure of bacteriophage phiCb5 reveals a role of the RNA genome and metal ions in particle stability and assembly. *J. Mol. Biol.* **391**:635–647.
14. Ruokoranta, T. M., A. M. Grahn, J. J. Ravantti, M. M. Poranen, and D. H. Bamford. 2006. Complete genome sequence of the broad host range single-stranded RNA phage PRR1 places it in the *Levivirus* genus with characteristics shared with alleviviruses. *J. Virol.* **80**:9326–9330.
15. Schmidt, J. M. 1966. Observations on the adsorption of *Caulobacter* bacteriophages containing ribonucleic acid. *J. Gen. Microbiol.* **45**:347–353.
16. Schmidt, J. M., and R. Y. Stanier. 1965. Isolation and characterization of bacteriophages active against stalked bacteria. *J. Gen. Microbiol.* **39**:95–107.
17. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
18. Zamora, H., R. Luce, and C. K. Biebricher. 1995. Design of artificial short-chained RNA species that are replicated by Q β replicase. *Biochemistry* **34**:1261–1266.