Bacteriophages with ssRNA

Secondary article

Jan van Duin, *Leiden University, Leiden, The Netherlands*

The single-stranded RNA (ssRNA) phages include about 100 isolates. The best-characterized are those infecting *Escherichia coli* , a Gram-negative bacterium found in the gut and faecal sewage. The genome of ssRNA phages assumes a highly optimized and unique structure that regulates translation of genes, confers RNAase resistance in the bacterial cytoplasm and ensures recognition by replicase.

Article Contents

Introduction

Since their discovery in 1961 by Loeb and Zinder, the ribonucleic acid (RNA) phages have served as a model system for the investigation of a variety of problems in molecular biology. As a source of homogeneous and readily obtainable messenger RNA, they have been particularly helpful in solving questions on the regulation of gene expression at the level of translation. The concepts of translational polarity and translational control by repressor proteins resulted from early studies on bacteriophage RNA.

The RNA phages have also provided us with the best-defined RNA replication system, currently used to study template recognition, *in vitro* RNA recombination and replication. Available infectious clones have allowed the investigation of the compromise between the need to fold the RNA in a specific way and the need to encode proteins.

Classification

The single-stranded RNA (ssRNA) phages include about 100 isolates. The best-characterized are those infecting *Escherichia coli*, a Gram-negative bacterium found in the gut and faecal sewage. Several coliphages, in particular MS2 and Qβ, have been the subject of intensive research and the RNAs of about a dozen of them have been fully sequenced. ssRNA phages have also been found in the Gram-negative bacteria *Acinetobacter* (AP205), *Caulobacter crescentus* (ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, others) and *Pseudomonas aeruginosa* (PP7 and 7s).

All ssRNA phages infect their hosts via pili. *Caulobacter* and *Pseudomonas* phages use polar pili, but in *E. coli* the sex or F-pili expressed by male (F^+) bacteria are used. The strong dependence on pili is exemplified by PRR1, a phage that infects many different bacteria such as *E. coli*, *P. aeruginosa*, *Salmonella typhimurium* and *Vibrio cholerae*, provided that they express the pili encoded in drug-resistance factor RP (P incompatibility group, e.g. RP1, RP4 or R1822). Similarly, members of several enterobacterial genera, such as *Shigella*, *Proteus* and *Salmonella*, can be artificially converted to coliphage sensitivity by introducing the F-factor of *E. coli* (expressing the F-pili) in these bacteria.

The ssRNA phages are members of the *Leviviridae* family, which presently includes the two genera *Levivirus* and *Allolevivirus* (**Table 1**). The first genus, also known as group A, is divided into subgroups I and II, whereas the alloleviviruses (group B) comprise subgroups III and IV. These subgroups have presently acquired species status. The differences between genera and species are detailed in the section Genome organization.

The taxonomy of these phages is based on morphology, serotyping, size of nucleic acid and structural proteins, and

Table 1 Taxonomic structure of the *Leviviridae* family

Host	Genus	Species	Members
Escherichia	*Levivirus* (group A)	I	MS2[a], fr[a], JP501[a], M12[a], R17[b]
		II	GA[a], BZ13, JP34[b], JP500, KU1[a]
	Allolevivirus (group B)	III	Qβ[a], M11[a], MX1[a], TW18[b], VK[b]
		IV	SP[a], FI, ID2, NL95[a], TW19
Acinetobacter	?	?	AP205[b]
Caulobacter	?	?	ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r
Pseudomonas	*Levivirus* (PP7)	?	PP7[a], 7s
Polyvalent	?	?	PRR1

[a] RNA completely sequenced.
[b] RNA partially sequenced.

ultimately the genomic map. Except for *Pseudomonas* phage PP7, most of these parameters are unknown in noncoliphages and the classification of the latter is not advanced. Most of the phages mentioned here are kept in the Watanabe Collection at Keio University, Tokyo, Japan, in the Félix d'Hérelle Reference Centre for Bacterial Viruses at Laval University, Quebec, Canada, or in the laboratory of the author.

Isolation and Identification

ssRNA phages depend on host pili to enter the bacterium. In *E. coli* the F- or sex pili, which are expressed by the male bacteria, are used. Thus RNA coliphages can be isolated by plating samples on F^+ cells. When RNAase is added, no plaques are formed. Coliphages can be identified either immunologically or by labelled DNA probes complementary to constant sequences present in the (sub)groups.

Virion Properties

Morphology/structure

In addition to one molecule of positive-strand RNA, each virion in group A contains 180 copies of the coat protein and one copy of the maturation, or A, protein. The group B phages contain in addition about 12 copies of the read-through protein in their capsid. For these reasons, the protein shell of the ssRNA phages is not isomeric like other icosahedral viruses such as poliovirus or satellite tobacco necrosis virus. The diameter of the phage is 26 nm, and the protein shell is about 2 nm wide (**Figure 1**). The nonenveloped quasi-icosahedral shell has a $T = 3$ surface lattice. Crystals of phage MS2, GA, fr and Qβ have been obtained. The coat protein structure of MS2 has been solved to 0.27 nm resolution by X-ray diffraction. Unfortunately, the RNA and the A protein are not seen in the electron density map. These molecules are apparently not ordered in the crystal.

Physical and chemical properties

The sedimentation constant ($s_{20,w}$) of the coliphages ranges from 79 to 84S. Buoyant density in CsCl is between 1.44 and 1.47 g cm^{-3}. Infectivity is ether- and chloroform-resistant. Temporal exposure to pH 2 is withstood.

RNA

The genome size within each species is remarkably constant. The genome of species II is shortest with 3466 nucleotides (GA) followed by those of species I (MS2) with 3569, species III (Qβ) with 4217 and species IV (SP) with

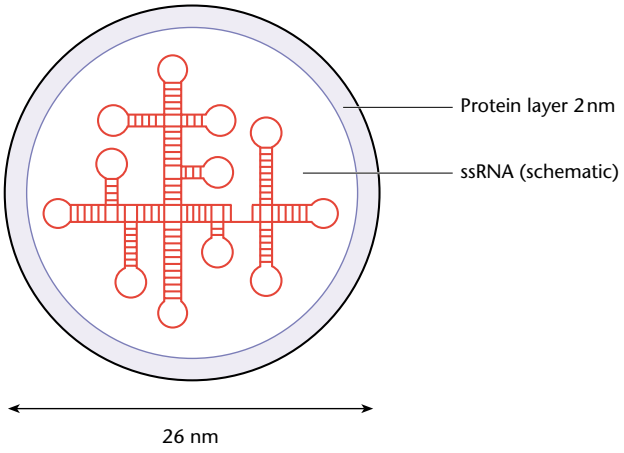

Figure 1 Schematic view of an RNA bacteriophage. The RNA is highly ordered and has very few unpaired regions.

4276 nucleotides, respectively. PP7, a levivirus growing in *Pseudomonas*, has 3588 nucleotides. All ssRNA phages start with 5′ pppGG and end with CCA_{OH}, where the terminal A is a noncoded base added by the terminating replicase. All four bases occur at about equal ratios in each phage RNA.

The RNA in each group assumes a specific secondary structure that is essential for the viability of the phage while it facilitates strand separation, confers resistance to host RNAases, regulates translation, provides coat protein-binding site(s) and creates a surface that is recognized by replicase as a replicable RNA. Besides these properties there may be functions for the structure that we do not yet know of. A further possibility is that the final shape of the RNA facilitates packaging.

The importance of a specific local or long-distance base pairing can be easily assessed by introducing neutral substitutions that break up the analysed structure. Depending on the particular role of the element, such mutations can decrease the plaque-forming potential over orders of magnitude.

Genome Organization

The genomic map of the two genera *Levivirus* and *Allolevivirus* is shown in **Figure 2**. The conspicuous difference between groups A and B is the presence of a read-through protein in group B. Here, the major coat protein gene ends in a leaky UGA stop codon that is read as tryptophan (UGG) with a probability of about 5%. This leads to a C-terminally extended coat protein which is incorporated in the capsid and which is required for infectivity. The other major divergence is the presence of a separate lysis gene in group A. This out-of-frame overlapping gene encodes a roughly 70-amino-acid long

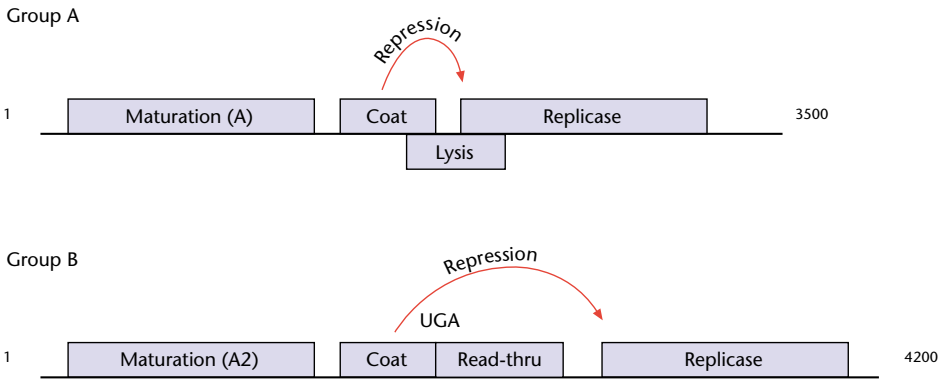

Figure 2 Genetic maps of group A and B RNA coliphages. Repression of replicase translation by the coat protein is shown. Lysis overlaps coat and replicase genes in the +1 frame.

hydrophobic peptide that short circuits the cytoplasmic membrane of the bacterium. Somehow, loss of membrane potential distorts the balance between components of the bacterial enzyme ensemble that cleaves and extends the peptidoglycan network. This leads to cell lysis. Group B phages do not encode a lysis peptide. Instead, the maturation protein has a dual function; it is a constituent of the virion and it triggers cell lysis. Accordingly, bacteria that overproduce the Qβ maturation protein lyse, while Qβ mutants carrying an amber mutation in the maturation protein gene do not cause bacteriolysis.

Genetic differences within the genera are subtle. In group A the most pronounced one is a 60 nucleotide insertion in the 3′ untranslated region of subgroup I. A 90 nucleotide insertion in the A2 protein gene of subgroup IV represents the major difference with subgroup III.

Replication

RNA entering the cell serves as messenger for the synthesis of the phage-coded replicase subunit. After assembly with several host proteins the minus strand can be synthesized, forming the template for new plus strands. As one phage RNA is sufficient to infect the cell, the specificity of the enzyme must be great.

The quicker *E. coli* grows, the better the infection proceeds. Several thousands of phages can be produced within a few hours from a single parent.

Host–virus relationship

The RNA gains access to the interior of the cell via long, filamentous structures called pili. These can be of various origins and serotypes. In *E. coli*, the sex (or F) pili are employed as vehicles (**Figure 3**) but in *Pseudomonas* and *Caulobacter* polar pili are used for this purpose. Transfer of RNA via pili is not essential for infection, however. Cells that lack pili can be infected by naked RNA if they are converted to spheroplasts.

The attachment of the phage to the sides of the pili proceeds via the maturation (A) protein. The coat protein is dispensable for this process; bacteria can be infected with a binary complex of A protein attached to phage RNA. In group B phages the minimal infection set requires the additional presence of the coat read-through protein.

Contact of phage MS2 with the pilus results in cleavage of its A protein into a 15-kDa and a 24-kDa fragment. This cleavage probably triggers the ordered ejection of the tightly packed RNA from the virus shell. The binding sites of the MS2 A protein on the RNA have been determined at nucleotide regions 400 and 3500, respectively. When cleavage of the A protein occurs between the two RNA-binding domains of the protein this would potentially lead

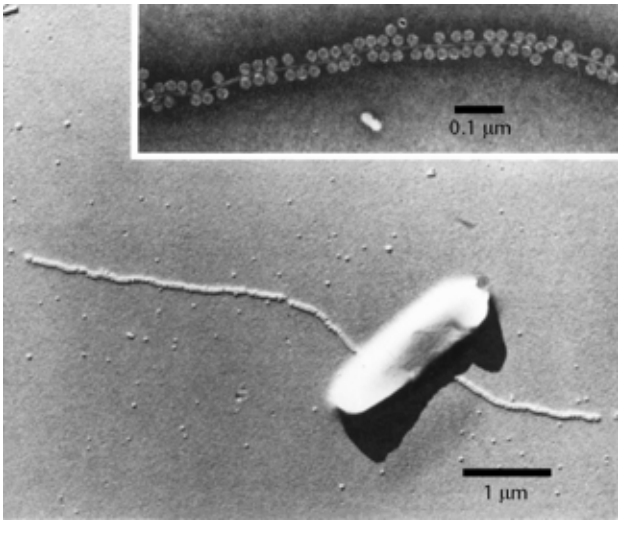

Figure 3 RNA phages attached to the F-pili of *Escherichia coli*. From Zinder ND (ed.) (1975) *RNA Phages*. New York: Cold Spring Harbor Laboratory. Contribution of W. Paranchych.

to the liberation of the two ends. Possibly, the 5′ end of the RNA begins to move along the pilus towards the cell. This stage of infection corresponds to the RNAase-sensitive step. The two A protein fragments remain associated with the RNA during penetration of the cell envelope. However, it is unlikely that these protein fragments play any further role, since the naked RNA is fully able to generate infectious progeny in spheroplasts. Also, transformation of *E. coli* with plasmids containing phage complementary DNA (cDNA) leads to productive infection.

RNA replication

Most of our knowledge on replication has been obtained in the Qβ system, but it is assumed that the principles also apply to the other groups. The replicase holoenzyme contains four different proteins called subunits I to IV, or subunits α, β, γ and δ. Subunit I was identified as ribosomal protein S1, and subunits III and IV are the translation elongation factors EF-Tu and EF-Ts. Thus, three proteins that normally function in the synthesis of proteins are recruited by the phage to assist in RNA synthesis. Subunit II is encoded in the viral genome, and its function is probably limited to the polymerization process. The four subunits occur in the enzyme complex in one copy. Replication proceeds via the synthesis of a free negative strand. Annealed positive and negative strands are not a substrate for the replicase enzyme.

The requirements for copying the positive and the negative strand are different. To copy the negative strand, only subunits II, III and IV are required, whereas positive strand replication also needs subunit I. In addition, copying the positive strand is greatly dependent on the product of the bacterial *hfq* gene, called host factor (HF). In the uninfected cell one function of this protein seems to be to promote the accessibility of highly structured messengers such as rpoS messenger RNA (mRNA) to ribosomes. In replication, its role may be to promote the accessibility of the 3′ end. In the absence of host factor, replication of phage RNA starts only if the base pairing occluding the 3′ terminus of Qβ RNA is destabilized or if a few unpaired C residues are added to the 3′ end. Group A uses a different host factor for copying the positive strand, but the protein has not yet been identified.

Replicase binds to two internal Qβ RNA sequences, called the S and M sites. This binding is mediated by the relatively unspecific protein S1. The interaction with the S site, overlapping the start of the coat gene, provides for the necessary competition with the ribosome, but it is not required for replication itself. Interaction with the M site, some 1500 nucleotides away from the 3′ end, is essential. It is supposed that in this RNA–replicase complex the folding of the RNA places the 3′ terminus in the active site of the enzyme. In this concept then, the specificity of the enzyme is derived from the shape of the RNA and not from sequences. From this point of view it is easy to see that Qβ replicase can multiply group IV RNAs as the RNA foldings in groups III and IV are very similar. But the enzyme cannot replicate group A RNA, because it has a very different RNA secondary structure.

Replication of the negative strand depends on structure elements at both the 5′ and 3′ termini and is thus basically different from positive strand copying. Here, the interaction between template and α-less replicase is reportedly mediated by EF-Tu.

Gene expression

The appearance of the phage-coded proteins in the infected cell is carefully controlled in time and quantity. For instance, replicase is a minor early product, while the amount of coat protein exceeds by far that of the other proteins. Since no DNA intermediates occur in the life cycle of the phage, control is predominantly exerted at the translational level: it is the RNA secondary structure that regulates access of ribosomes to initiation regions.

In MS2, the maturation protein is needed at only one copy per virion, and its translation is accordingly kept at a low level. It was found that it can only be translated from a nascent strand. Its ribosomal-binding site (RBS) is base paired to an upstream sequence and is accessible during a short period when the growing RNA has not yet reached its equilibrium folding. In the equilibrium structure the RBS is inaccessible. For Qβ the RNA structure model predicts a subtly different mechanism. Also here, the RBS of the maturation gene is inaccessible in full-length Qβ. However, the time window is created by the fact that the RBS is base paired to a downstream sequence. As a consequence all nascent RNA that does not yet contain the complementary downstream sequence can be translated. A completely different model suggesting downregulation by coat gene translation has been published recently for Qβ.

In group A, both replicase (R) and lysis (L) genes are under translational control of the coat gene, as witnessed by the observation that nonsense mutations at the beginning of the coat gene inhibit expression of the R and L genes. For the replicase the underlying mechanism is base pairing between the start of the R gene and a coding region of the coat gene. Coat protein-synthesizing ribosomes pass through this region, destroy the pairing and temporarily liberate the R start. Coupling of the lysis gene arises by a local hairpin burying the L start. Exposure is brought about by a ribosome that has arrived at the coat gene terminator codon. After releasing the synthesized coat protein this ribosome can, albeit with a low efficiency, reach the L start site by random movements along the RNA. This process resembles the scanning in eukaryotes, whereby ribosomes are thought to move along the RNA from the 5′ cap to the first start codon.

The real downregulator of replicase is the coat protein. Once present in sufficient concentration, coat protein dimers bind to a hairpin structure that contains the R-start, thereby preventing any ribosome binding. The complex between the hairpin and the coat protein dimer has been modelled by X-ray analysis.

Translational coupling of L and R genes, and the temporary expression of the A gene are not only simple ways to cut down on the product levels, but the designs also serve a more sophisticated purpose, in that on full-sized RNA the only site independently accessible to ribosomes is the start of the coat gene. As discussed above, replicase binding also involves this start site. The resulting competition between ribosome and replicase means that the RNA is either used for replication or for translation. The importance of these controls is further illustrated by the finding that upon their distortion by targeted mutagenesis there is rapid selection of pseudorevertants in which regulation is reinstalled.

Apart from the occasional competition by the replicase, the coat protein RBS is not negatively regulated. In fact, this site is positively affected by the upstream sequence. It is not known how this happens, but the presence of the upstream 1300 nucleotides stimulates translation 30-fold.

Particle assembly

Capsid formation can, in principle, occur in the absence of the viral RNA. In the presence of the RNA a much lower coat protein concentration suffices to build capsids. This is a result of the binding of a coat protein dimer to the replicase start hairpin which then acts as a nucleation site. Without the maturation protein, however, the particles are not infectious.

It is believed that binding of the A protein to MS2 RNA is an early assembly event, preceding capsid formation.

Ecology

Furuse has examined and reviewed the geographical distribution of the single-stranded RNA phages as well as their present-day habitat. They are most frequently encountered in sewage and faeces of mammals, and their titres in sewage samples may be as high as 10^7 plaque-forming units (PFU) mL^{-1}. RNA phages may constitute up to 90% of total coliphages present in these samples, but the number can vary substantially.

The geographical distribution of groups II and III shows a strong bias. In northern Japan, there is a relative abundance of group II over group III (6:1) per sampling site. Moving southward, this ratio drops dramatically until in Southeast Asia group II becomes rare. Furuse has suggested that the north–south gradient is related to differences in climate. Group III (and also I and IV) propagates well at 40°C but not at 20°C, whereas for group II the situation is reversed. A problem here seems that group II is rare in some countries e.g. The Netherlands, where the average temperature is below 20°C.

Although the natural host for the single-stranded RNA phages is not known with certainty, it is clear that they propagate stably in the intestines if male *E. coli* is present as host. Thus *E. coli* can sustain the life cycle of the phage under 'natural' circumstances. In Japan and The Netherlands attempts have been made to determine whether certain phage (sub)groups are preferentially associated with certain animal species or with humans. So far, these studies have not been conclusive.

Evolution

Two kinds of evolution for organisms can be distinguished. One, the common one, describes and comprises all that has happened genetically to the organism since it came into being. For the RNA phages it is generally assumed that they all derive from a common ancestor because of the nearly identical genetic organization, the strong resemblance of the replicases, the use of the same host proteins (except host factor) as auxiliaries in the copying reaction, and the similarities of several control mechanisms. These properties are more easily explained by divergent than by convergent evolution.

The second kind of evolution refers to the response of organisms to artificial sequence changes in their genes. In particular in RNA viruses with their high mutation rate, their large number of offspring per infection cycle and their short generation time, the selection and perfection of pseudorevertants, adapted to the introduced base changes, can be completed within 24 h. It is a powerful technique that allows the existence and importance of base-paired structures to be assessed. It has provided insight into the flexibility of the genome in absorbing extra nucleotides, and the extent to which the RNA structure can be rearranged while maintaining phage viability. The pseudorevertants obtained in this way are not easily distinguished from wild type; they produce the same numbers of plaques and cause cell lysis to the same extent. However, in a direct competition they cannot stand up to wild type.

Rigidity and plasticity of the genome structure

Considering the inaccuracy of phage RNA replication, one might expect the sequence of a phage to drift away or to become heterogeneous. This turns out not to be true. Despite many years of laboratory cultivation, the sequences of Qβ and MS2 have hardly changed. Similarly, it has turned out to be quite difficult to find RNA phages in nature whose sequences diverge substantially from the subgroup prototype. For instance, species I representatives like MS2, f2, R17, M12 and JP501, all isolated independently in different parts of the world, show more than 95%

sequence identity. Thus there seem to be very few solutions that are good enough to coexist. It is assumed that the selection pressure, which discards all potential variants (for instance those having synonymous codons), originates from the demands on the RNA secondary structure for phage fitness. As mentioned, such contributions of structure involve regulatory circuits, RNAase resistance, preventing the annealing of positive and negative strands and probably many other parameters.

In spite of this apparent rigidity, the genome structure can show a high degree of flexibility under noncompetitive conditions. As mentioned, laboratory evolution of phages containing a wanton distortion in RNA structure yields many pseudorevertants presenting us with a large spectrum of structural solutions. These solutions all perform well as long as they do not have to compete with their wild-type counterpart.

Applications

RNA phages have the same morphology and chemical properties as many pathogenic enteroviruses. Thus they can serve as model viruses to study the effectiveness of water treatments such as ultraviolet irradiation, chlorination or ozonization.

In addition, as enteroviruses and RNA phages share the same habitat, the gastrointestinal tract, they can serve as indicator organisms for the presence of pathogens in drink and surface water. RNA phages cannot multiply in surface water since the temperature is generally too low for the formation of F-pili by *E. coli*. DNA probes have been developed that can identify plaques as RNA phages.

There is presently renewed interest in DNA phages as combattants of bacterial infections. A role for RNA phages is doubtful. Early pilot studies showed that *E. coli* adapts to RNA phages by losing its F-pili.

Further Reading

Armon R and Kott Y (1996) Bacteriophages as indicators of pollution. *Critical Reviews in Environmental Science and Technology* **26**: 299–306.

Furuse K (1987) Distribution of coliphages in the environment: general considerations. In: Goyal SM (ed.) *Phage Ecology*, pp. 87–124. New York: Wiley.

Havelaar HA (1991) Bacteriophages as model viruses in water quality control. IAWPRC Study Group on Health Related Water Microbiology. *Water Research* **25**: 529–545.

Olsthoorn RCL, Licis N and van Duin J (1994) Leeway and constraints in the forced evolution of a regulatory RNA helix. *EMBO Journal* **13**: 2660–2668.

Valegård K, Murray JB, Stockley PG, Stonehouse NJ and Liljas L (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* **371**: 623–626.

Van Duin J (1988) In: Fraenkel-Conrat H and Wagner RR (eds) *The Bacteriophages*. Series *The Viruses*, pp. 117–167. New York: Plenum Press.