

Chapter 13

Protein-RNA Interactions in the Single-Stranded RNA Bacteriophages



Jānis Rūmnieks and Kaspars Tārs

Introduction to ssRNA Phages

The bacteriophages with single-stranded RNA (ssRNA) genomes are among the simplest and smallest of the known viruses. Due to their simplicity, these phages have for long been used as models to study fundamental problems in molecular biology, such as translational control mechanisms, protein-RNA interactions, RNA replication, virus evolution, structure, and assembly. In 1976, the ssRNA phage MS2 was the first life form for which the complete genome sequence was determined (Fiers et al. 1976). The ssRNA phages have also been the source for many diverse applications, including vaccine development, imaging tools, and ecological and virus inactivation studies (see Pumpens et al. 2016 for a review).

All of the known ssRNA phages belong to the *Leviviridae* family and have small, approximately 3500 to 4200 nucleotide long genomes that encode just a few proteins (Fig. 13.1). Three of the proteins – the maturation, coat, and replicase - are conserved among all ssRNA phages. Many of the studied phages also have a short open reading frame (ORF) that codes for a lysis protein, which often overlaps with other genes and shows a surprising variation in its location within the genome (Klovins et al. 2002; Kazaks et al. 2011; Rumnieks and Tars 2012). Distantly related lysis proteins lack any sensible sequence identity and have presumably arisen several times independently from each other. All of the lysis polypeptides have one or two predicted transmembrane helices and are thought to cause cell lysis by forming ion-permeable pores in the periplasmic membrane (Goessens et al. 1988), which leads to depolarization of the membrane and subsequent activation of autolysins that degrade the cell wall. A subgroup of the ssRNA phages that are assigned

J. Rūmnieks · K. Tārs (✉)
Biomedical Research and Study Center, Riga, Latvia
e-mail: j.rumnieks@biomed.lu.lv; kaspars@biomed.lu.lv

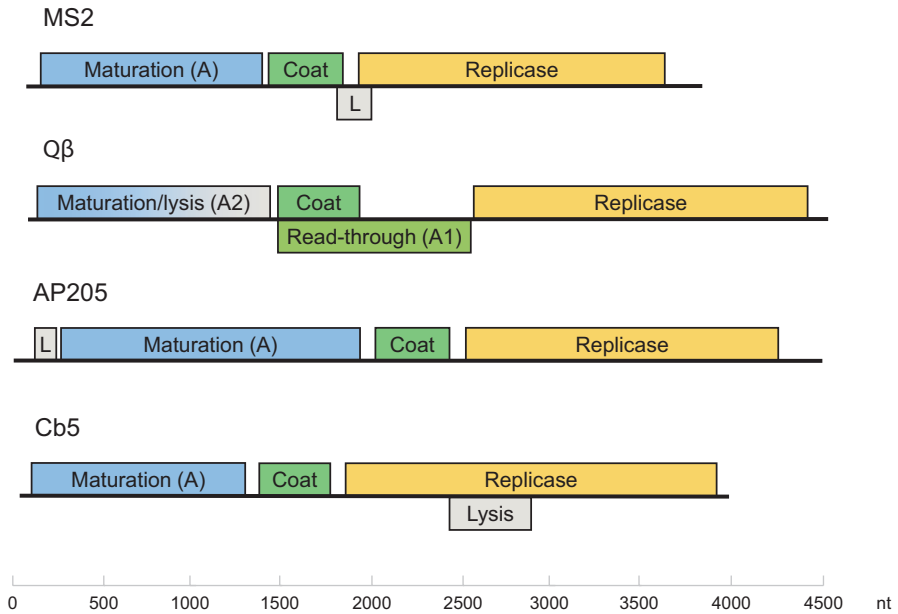


Fig. 13.1 Genome organization of the single-stranded RNA bacteriophages. The genes are represented as boxes; L, lysis gene. In bacteriophage Q β , the maturation protein mediates cell lysis and A1 is a C-terminally extended variant of the coat protein generated by ribosomal read-through of the coat gene

to a separate *Allolevivirus* genus does not have a dedicated lysis protein; instead, the cell lysis is accomplished by the maturation protein which blocks an enzyme in the peptidoglycan biosynthesis pathway (Karnik and Billeter 1983; Winter and Gold 1983; Bernhardt et al. 2001). Another distinctive feature of the alloleviviruses is the presence of the so-called A1 protein in the capsid, which is an elongated version of the coat protein produced by a translational read-through mechanism (Weiner and Weber 1971). The exact function of the A1 protein is unknown, but it is important for the infectivity of the particles (Hofstetter et al. 1974). The only other recognized genus in the family, *Levivirus*, includes several phages where the lysis ORF is in a position overlapping the coat and replicase genes. However, many other ssRNA phages currently remain unassigned to any genus, and recent metagenomic studies have unraveled a vast array of new ssRNA phage sequences, in many cases very distantly related to the currently described ones (Krishnamurthy et al. 2016; Shi et al. 2016). Therefore, it should be noted that as the true ssRNA phage diversity in nature begins to be realized, a future reclassification within the *Leviviridae* family appears imminent, with a possible dissolution of the currently recognized *Levivirus* and *Allolevivirus* genera.

Structurally, the *Leviviridae* virions are composed of a single RNA molecule packaged inside a protein shell that is approximately 28 nm in diameter and consists of 178 coat protein molecules and a single copy of the maturation protein (Fig. 13.2a).

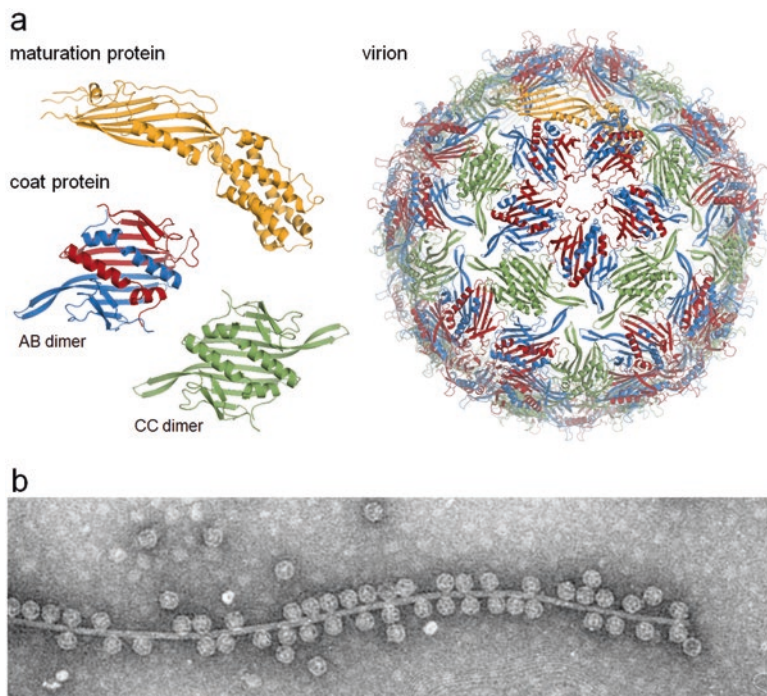


Fig. 13.2 Structure of an ssRNA bacteriophage particle. **(a)** Protein components of the virion. The coat protein dimers exist in two quasi-equivalent conformations in the particle, denoted AB (blue/red) and CC (green). An assembled particle consists of a single copy of the maturation protein, 60 coat protein dimers in the AB conformation and 29 in the CC conformation. The maturation protein replaces a single coat protein CC dimer in the otherwise icosahedrally symmetrical particle. **(b)** An electron micrograph of MS2 bacteriophage particles bound to an F pilus. Figs. 13.2a, 13.4, 13.5, 13.6, and 13.7 were prepared using Pymol version 1.8

The coat protein forms very stable dimers; therefore the capsid is more precisely described as composed of 89 coat protein dimers, and the maturation protein replaces a position otherwise occupied by a single coat protein dimer (Dent et al. 2013; Koning et al. 2016). The maturation protein serves as the attachment protein for the phage and mediates the adsorption of the virion to bacterial pili (Fig. 13.2b), which the ssRNA phages use as receptors for infecting the cell. The pili used by different ssRNA phages are rather distinct, ranging from the F-plasmid-encoded conjugative pili that the *E. coli* phages MS2 and Q β employ (Crawford and Gesteland 1964), to various genome-encoded pili used by *Pseudomonas* phage PP7 (Bradley 1966), *Acinetobacter* phage AP205 (Klovins et al. 2002), or *Caulobacter* phage Cb5 (Schmidt and Stanier 1965). After adsorption, the maturation protein leaves the capsid together with RNA and guides its entry into host cell via a poorly understood mechanism. A complex of only the maturation protein and the genomic RNA is infectious to the cell (Shiba and Miyake 1975), and the coat protein does not have another role than protecting the genome before the infection takes place.

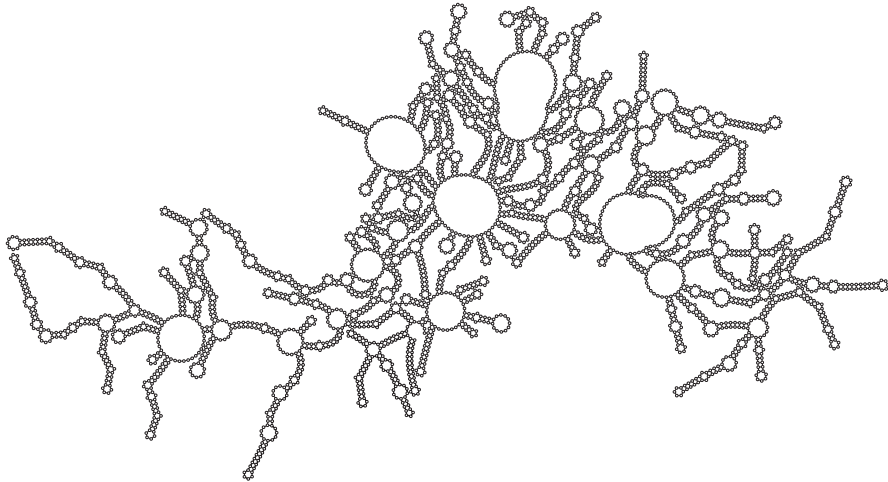


Fig. 13.3 Secondary structure of an ssRNA phage genome. The minimum free energy structure of the MS2 genomic RNA as predicted by the RNAfold software (Zuker and Stiegler 1981). The image was prepared using RNAfdl (Hecker et al. 2013)

The RNA genome is an equally important structural component of the ssRNA phages. While the notion that the genome of the ssRNA phages is a single-stranded RNA is correct, this is only true in the sense that each virus particle indeed contains a single RNA strand. However, as much as 75% of the RNA bases are involved in short- and long-distance base-pairing interactions within the genome (Skripkin et al. 1990) (Fig. 13.3), which renders most of the genome double stranded and results in a complex three-dimensional structure. The maturation, coat, and replicase proteins are all RNA-binding proteins that recognize specific RNA structures in the genome, and different protein-RNA interactions are of essential importance during the ssRNA phage life cycle.

Replicase-RNA Interactions

As cells do not contain an enzyme capable of synthesizing long RNA molecules from an RNA template, all RNA viruses have to supply their own RNA-dependent RNA polymerase (RdRp) for the purposes of replicating their genome. All ssRNA phages likewise encode a 60–65 kDa polypeptide with the enzymatic RdRp activity; however, the protein alone is not capable of replicating the genome. The phage-encoded protein, often referred to as the “ β subunit,” recruits three more proteins from the bacterial cell, the ribosomal protein S1 (Wahba et al. 1974), and translation elongation factors EF-Ts and EF-Tu (Blumenthal et al. 1972), that together assemble into the replicase holoenzyme complex. The normal function of EF-Tu in the cell is to bind amino-acyl tRNAs and deliver them to ribosomes, while EF-Ts acts as a

guanine nucleotide exchange factor for EF-Tu. The ribosomal protein S1 is a translation initiation factor that consists of six consecutive OB (oligonucleotide/oligosaccharide-binding) domains, of which the two N-terminal domains bind to the small ribosomal subunit, while the rest of the protein interacts with mRNA or auto-regulates its own synthesis (Boni et al. 2000). The S1 protein differs from the other subunits in that it is not required for the structural integrity of the replicase complex and a “core replicase” enzyme consisting of only the β subunit, EF-Tu and EF-Ts, is enzymatically active.

The discovery of several host-derived subunits in ssRNA phage replicases seemed rather surprising at first, but since then it has become clear that the ssRNA phages are hardly unique in this respect. The idea of borrowing and repurposing host RNA-binding proteins appears to be fairly popular also among many eukaryotic RNA viruses, which often recruit proteins from the host’s translation machinery to assemble a fully functional RdRp complex. Interestingly, several plant and animal viruses use the translational elongation factor eEF1A, a eukaryotic counterpart of EF-Tu, although not exactly in the same way as the ssRNA phages (see Li et al. 2013 for a review). Still, the function of the host-derived proteins in replication is probably the best understood in the ssRNA phage replicases.

Most of what is known about ssRNA phage RdRps, and their RNA replication in general, comes from studies of the enzyme from bacteriophage Q β . The structure of the Q β core replicase resembles a boat where the catalytic β subunit is located at one end and EF-Ts and EF-Tu at the other, with the active center facing the inner cavity of the structure (Kidmose et al. 2010; Takeshita and Tomita 2010) (Fig. 13.4). The β subunit has an architecture similar to other RdRps with the right-handed palm, thumb, and finger domains. EF-Tu participates in RNA binding during the elongation stage and forms part of the template exit channel, while the main function of EF-Ts appears to be the stabilization of the other subunits in an active conformation. The S1 protein binds to the opposite side of the β subunit with the same two N-terminal OB domains that are used for binding to the ribosome (Takeshita et al. 2014). The S1 protein is required to recognize and initiate replication of the genomic RNA strand (Kamen et al. 1972), and recently, the two protein-bound N-terminal S1 domains have been also proposed to function as a termination factor for an efficient release of product and template strands in a single-stranded form (Vasilyev et al. 2013).

The Q β replicase has a remarkable processivity and can generate up to 10^{10} copies of some in vitro selected templates in 10 min (Chetverina and Chetverin 1993). At the same time, the enzyme is strongly selective in which RNAs are replicated well, and the natural template, the Q β genome, is highly adapted to be efficiently replicated by the Q β replicase. As the first requirement, the template needs to be single stranded, and the enzyme cannot initiate synthesis on double-stranded RNA (Weissmann et al. 1967). In the phage genome, the large proportion of RNA secondary structures ensures that the (+) and (–) strands do not anneal during replication and remain separate. To initiate RNA synthesis, the Q β replicase does not require a primer but instead relies on a trinucleotide sequence CCA at the very 3′-terminus of the template (Chetverin and Spirin 1995). Intriguingly, tRNAs also have a CCA-3′

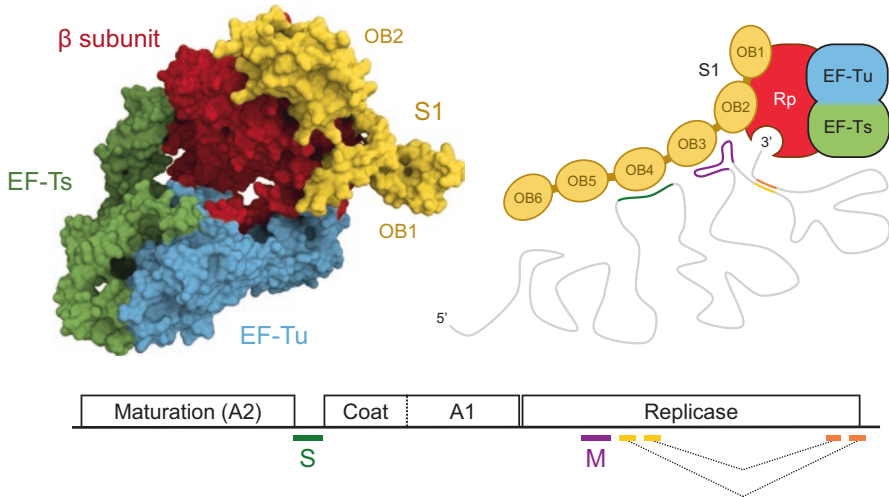


Fig. 13.4 RNA binding of the Q β replicase. Left, the three-dimensional structure of the replicase complex. The phage-encoded catalytic β subunit (red) recruits three proteins, elongation factors EF-Tu (blue) and EF-Ts (green), and ribosomal protein S1 (yellow) from the host that together assemble into a holoenzyme complex. Right, a model for the genomic RNA recognition by the Q β replicase. The S1 protein recognizes two internal sites in the Q β genome (bottom), the S site (green) and the M site (violet). Two long-distance base-pairing interactions (orange/yellow) bridge the 3'-untranslated region of the genome to a nucleotide stretch nearby the M site, which constrains the genome in a particular conformation. Consequently, binding of the replicase-bound S1 protein to the genome positions its 3'-terminus in the active center of the β subunit

sequence which is recognized by EF-Tu, and it was for once thought that the phage has hijacked this ability to recognize its own genome. However, high-resolution structures of the Q β replicase complex captured in the initiation stage have shown that the 3'-terminus is recognized solely by the β subunit where the 3'-terminal adenosine is kept in position by multiple contacts with the protein and by stacking interactions with the penultimate C and its complementary GTP (Takeshita and Tomita 2012). This way, the 3'-adenosine serves as a stable platform to initiate the replication, which begins at the penultimate cytidine, and not the adenosine itself. Upon termination, the Q β replicase adds a non-templated adenosine to the newly synthesized strand (Weber and Weissmann 1970). To ensure exponential amplification, the original template thus needs to begin with GG so that the complementary strand ends with CCA and is able to guide the synthesis of another (+) strand.

A high degree of secondary structure and a sequence 5'-GG...CCA-3', however, are not sufficient for an RNA molecule to serve as a good template for the Q β replicase, and the recognition of the phage genome is considerably more complex. The Q β genome has two internal sites, the S site and the M site, which are recognized by the replicase holoenzyme (Meyer et al. 1981). The S site is an approximately 100-nucleotide-long uridine-rich stretch preceding the coat protein gene that is recognized by the S1 protein (Miranda et al. 1997). The S site is dispensable for

replication but serves a role in coordinating the replication and translation of the genome. The coat protein gene does not have a canonical Shine-Dalgarno sequence to initiate translation and requires the ribosome-bound S1 protein to recruit the ribosome. This creates a situation where the S1 proteins from both the ribosome and the replicase complex compete for the same site in the genome. In the folded phage genome, the only translation initiation site that is available to ribosomes is that of the coat gene, while those of the other genes are buried in secondary structure and inaccessible (Van Duin and Tsareva 2006). Therefore, binding of the replicase complex to the S site prevents ribosomes from translating the genome and grants it exclusive rights for of the (+) strand, which is of particular importance early in infection when it is much more beneficial to actively replicate the genome than translate a few existing copies. In addition, since the replicase has to constantly compete for the (+) strand whereas the complementary strands are always available for copying, the initiation rate on (–) strands is higher, which results in a favorable ten fold excess of the (+) strands (Chetverin and Spirin 1995). The M site is an about 100-nucleotide-long branched stem-loop structure (Schuppli et al. 1998) within the replicase-coding sequence that is part of a bigger RNA structural domain called RD1. The M site is also recognized by the S1 protein (Miranda, Schuppli et al. 1997), but in contrast to the S site, its removal results in a drastic loss of template activity (Schuppli et al. 1998). The M site and the 3′-terminus are more than a thousand nucleotides apart from each other but are brought in close vicinity by long-distance base-pairing that involves the 3′-untranslated region of the genome and a nucleotide stretch adjacent to RD1 (Klovins et al. 1998; Klovins and van Duin 1999). The S site is also apparently close to the M site in the folded genome, as both sites can be bound simultaneously by the S1 protein; the M site is bound primarily by the third OB domain (Takeshita et al. 2014), while the S site is bound most likely by the adjacent C-terminal OB domains. Thus, while the recognition of the genome by the replicase complex is arguably the most extensive of the known protein-RNA interactions in the ssRNA phages, the phage achieves this by a clever recruitment of the cellular S1 protein to exploit its RNA-binding capabilities. Binding of the replicase-constituent S1 protein to the M site apparently positions the complex in such a way that the 3′-terminus of the genome is brought into the active center of the β subunit which allows the RNA synthesis to be initiated (Fig. 13.4).

Another *E.coli* protein, “the host factor for Q β ” (hfq), has been identified that further enhances the replication of Q β RNA (Franze de Fernandez et al. 1968; Franze de Fernandez et al. 1972). Hfq is an abundant RNA-binding protein with several functions in the cell and directly binds to Q β genome, where it presumably further increases the availability of the 3′-terminus to the replicase (Van Duin and Tsareva 2006). However, in contrast to the other bacterial proteins that the phage makes use of, the host factor is not essential for Q β replication, and the phage can quickly adapt to grow without hfq by accumulating a few mutations in the genome (Schuppli et al. 1997; Schuppli et al. 2000).

The ssRNA phage replicases are among the most error-prone polymerases known, resulting in highly divergent sequences. Although much of what is learned from the Q β replicase likely applies to other ssRNA phages, the differences in

genome and protein structure surely have an impact. This appears to hold true for even reasonably closely related phages, such as Q β and another *E.coli* phage MS2. Although the MS2 genome contains both the 5'-CC and CCA-3' sequences and the S and M sites, it is not recognized as a template by the Q β replicase, and vice versa, the MS2 enzyme does not copy Q β RNA (Haruna and Spiegelman 1965). The discrimination is likely caused by differences in the complex three-dimensional structure of the two genomes, which are expected to only become more significant in increasingly distant ssRNA phages. Despite recent advances in the structural studies of the ssRNA phages, the molecular details of how the replicase binds to the whole genome are currently still unknown and await further investigations.

While the Q β replicase is highly adapted to recognize and replicate the phage genome, there do exist other RNA molecules that the enzyme is capable of replicating. In Q β -infected cells, a variety of shorter RNA molecules can be detected that had been derived from the Q β genome, in some cases by recombination with cellular RNAs (Munishkin et al. 1988; Munishkin et al. 1991; Moody et al. 1994; Avota et al. 1998). These shorter RNAs do not seem to have any biological function and are considered as mere by-products of the replicase activity. As shorter RNAs take less time to replicate, they gradually outcompete the longer phage genome, which in the limited time of infection apparently does not cause issues for the phage but are nevertheless interesting to study in vitro. In the most famous experiment, the Q β replicase was initially allowed to replicate the Q β genome in vitro, the reaction products transferred to another tube with fresh Q β replicase and nucleotides but no template, and the transfer was then repeated many times over. After 74 generations, an RNA molecule just 218 nucleotides, dubbed the "Spiegelman's monster," had emerged and was being replicated much better than the original phage genome (Kacian et al. 1972). Since then, many other artificial RNAs replicable by the Q β replicase have been described (Van Duin and Tsareva 2006). Like the phage genome, these RNAs have the expected 5'-GG...CCA-3' sequence and significant amounts of secondary structure, and apparently some kind of tertiary structure that has selected them as better templates than others. Some in vitro experiments had also suggested that the Q β replicase can generate RNA spontaneously without any template, given only a mixture of nucleotides (Sumper and Luce 1975; Biebricher et al. 1986; Biebricher and Luce 1993). However, later experiments have arrived at a general conclusion that the sometimes-observed template-free de novo RNA synthesis is an artifact caused by minute amounts of contaminating RNA, present in enzyme preparations, buffer solutions, labware, or even laboratory air (Chetverin et al. 1991).

Coat Protein: RNA Interactions

Repression of the Replicase Gene

The obvious function for the ssRNA phage coat protein is the formation of a protein shell that protects the genome during the extracellular stage of the phage life cycle. Yet another role for the coat protein, at least in a subgroup of the ssRNA phages, is

to regulate the translation of the replicase gene. The replicase is a characteristic early gene product that is required at the beginning of the infection, while later, when large amounts of phage RNA have been generated, it becomes beneficial to cease replication and switch to packaging of the genomes in new virus particles. In phage infected-cells, an observation can be made that as the amount of the synthesized coat protein grows, the synthesis of the replicase enzyme correspondingly diminishes. Behind the regulatory mechanism is the specific binding of the coat protein to an RNA hairpin structure at the very beginning of the replicase gene, usually referred to as the “translational operator” or “translational repressor” (Gralla et al. 1974). The hairpin contains the initiation codon of the replicase gene, which upon binding to the coat protein becomes masked from ribosomes, which in turn downregulates the translation of the replicase gene (Weber 1976).

The coat protein RNA operator interaction in the ssRNA phage MS2 has been extensively studied genetically, biochemically, and structurally and is one of the best understood protein-RNA interactions to date. The coat protein consists of an N-terminal β -hairpin, a five-stranded β -sheet, and two C-terminal α -helices. In a coat protein dimer, the two monomers form a continuous ten-stranded β -sheet that in the assembled particle lines the interior of the capsid and forms the RNA-binding surface of the protein (Valegård et al. 1990). The MS2 operator is a 19-nucleotide long RNA hairpin, composed of a seven-base pair stem with a single unpaired adenosine, and a four-nucleotide long loop. While the coat protein dimer is itself symmetric, the RNA operator binds asymmetrically across the RNA-binding surface (Fig. 13.5a), and each of the coat protein monomers interacts differently with the RNA. In the MS2 operator, four nucleotides, A-10, A-7, U-5, and A-4 (the base numbering is relative to the adenosine in the replicase initiation codon), contribute to the specific binding, but the crucial determinants for the interaction are the unpaired A-10 adenosine in the hairpin stem and the A-4 in the loop, which dock into two adenine-recognition pockets in the coat protein dimer (Valegård et al. 1994). Both pockets are identical, each formed by one of the coat protein monomers, but the interaction with the adenine bases is different in each pocket (Fig. 13.5b, c). The protein-RNA interaction is further stabilized by continuous aromatic stacking that via the A-7 and U-5 bases in the loop extends from the RNA stem to a tyrosine side chain in the coat protein. The sugar-phosphate backbone also makes extensive sequence-nonspecific interactions with the protein.

A lot of effort has been put toward characterizing many different MS2 operator variants to determine the exact contribution of the RNA bases for the binding interaction. For example, substitution of the U-5 base in the hairpin loop with a cytidine increases the coat protein-RNA affinity about 50-fold (Lowary and Uhlenbeck 1987). A crystal structure of the “C-variant” complex revealed that while the coat protein-operator interactions are essentially identical to the wild-type, the C-5 forms an additional intramolecular hydrogen bond in the RNA that stabilizes the operator structure and is apparently responsible for the tighter binding (Valegård et al. 1997). Several other structures with substitutions at the -5 position have been determined, with a general conclusion that the existence of the base stack itself is much more important than the identity of the bases (Grahn et al. 2001). In one case, a substitution

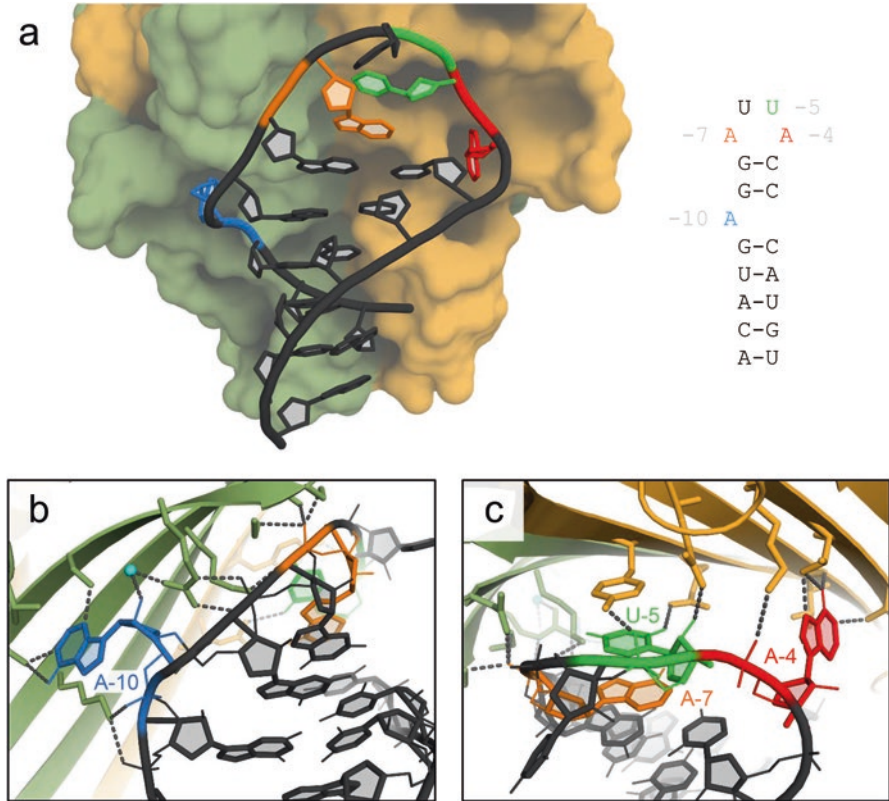


Fig. 13.5 RNA binding of the MS2 coat protein. (a) The overall RNA-binding mode. The coat protein dimer (light green/light orange as of the two monomers) binds to a hairpin structure (black) in the phage genome with high affinity. Four nucleotides (colored) are involved in the specific interaction. (b, c) A close-up view of the coat protein-RNA complex. Interactions around the unpaired adenosine in the hairpin stem (b) and the hairpin loop (c) are shown in the same colors as in (a)

of the -5 uracil with pyridin-4-one led to dramatic conformational rearrangements in the loop that caused the unnatural base to face away from the protein, but the stacking interaction with the tyrosine was still preserved by the neighboring U-6 base (Grahn et al. 2000). Substitutions at the -10 and -7 positions have likewise been tested, with a consensus that the loop is more important for binding than the unpaired base in the stem, but regardless the lower affinity, the coat protein is able to accommodate a wide variety of hairpin variants with only minor structural adjustments (Helgstrand et al. 2002).

The coat protein-RNA interactions have also been studied for several other ssRNA phages, albeit in much less detail compared to MS2. Bacteriophage PRR1 has an operator hairpin similar to MS2, except that it has a five- instead of a four-nucleotide loop. A crystal structure of the PRR1 coat protein-operator complex showed that the interaction, unsurprisingly, is almost identical to that of MS2

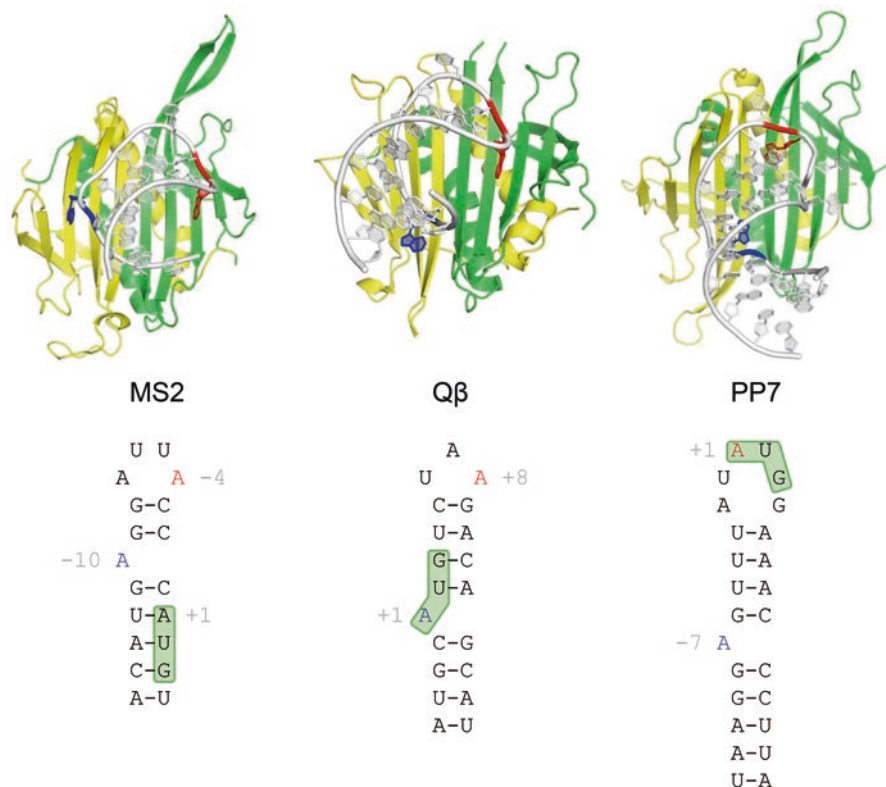


Fig. 13.6 Coat protein-RNA operator interactions in different ssRNA phages. Top, three-dimensional structures of phage MS2, Q β and PP7 coat protein-operator complexes. The protein is colored in green/yellow as of the two monomers, and the bound operator is shown in light gray. A bulged adenosine that is present in all of the operator stems is indicated in blue, and another adenosine in the hairpin loop that is important for the interaction in red. The corresponding operator sequences are presented below the structures. The numbering of the bases is relative to the initiation codon of the replicase gene (green box)

(Persson et al. 2013). The structure of the coat protein-RNA complex from bacteriophage Q β turned out to be more interesting (Rumnieks and Tars 2014). The Q β operator is rather different from that of MS2 with a three-nucleotide loop and the bulged adenosine at a different location relative to the loop (Fig. 13.6). Despite the differences, the overall RNA-binding mode of the MS2 and Q β coat proteins is similar, and the Q β adenine-binding pockets are almost identical to those of MS2. In the hairpin loop, the A + 8 in Q β and A-4 in MS2 operators make virtually identical interactions with one of the pockets, but the other adenine-binding pocket, which in MS2 is occupied by the bulged A-10, is empty in Q β . Instead, the unpaired A + 1 in the Q β operator makes a stacking interaction with a tyrosine residue, a mechanism that has not been observed in other ssRNA phages to date. The A + 8 adenine in the hairpin loop is the only sequence-specific requirement for the coat protein-RNA interaction in Q β , while the identity of the other nucleotides in the

loop, as well as the presence of the unpaired base in the stem, is dispensable for the interaction (Lim et al. 1996). Despite the seemingly lower specificity, the Q β coat protein is still able to discriminate in favor of its cognate operator and binds it with a comparable strength to the other studied ssRNA phages. For strong binding to the Q β coat protein, the RNA hairpin requires a three-nucleotide loop and an eight-base pair-long stem (Witherell and Uhlenbeck 1989). Compared to MS2, where virtually all of the protein-RNA contacts are located in the region between the bulged adenosine and the loop, in Q β a significant proportion of the interactions involve the lower part of the stem, which explains the greater length dependence for the RNA. Binding of a three-nucleotide loop with an adenosine in the 3'-most position orients the lower part of the RNA helix at a favorable position for interacting with the distant RNA-binding residues, while a four-nucleotide loop and an unpaired base at an MS2-like position would position the RNA stem differently, resulting in weaker binding.

The replicase operator in bacteriophage PP7 is markedly different from the other studied ssRNA phages, with a six-nucleotide loop and a bulged adenosine four nucleotides prior to the loop (Fig. 13.6). The interaction of the operator with the PP7 coat protein is also very distinct from that of the other phages (Chao et al. 2008). The PP7 interaction relies mostly on sequence-specific interactions between the RNA and the protein (Lim and Peabody 2002) and involves a total of four bases. Similarly to MS2, the unpaired adenosine in the PP7 operator stem and another one in the loop bind to two symmetrical adenine-recognition pockets on the coat protein surface. However, the pockets are unrelated to those of MS2 and are located in a completely different position on the dimer surface. Similar to other phages, also in the PP7 complex, two bases in the hairpin loop continue the base stack from the RNA stem. However, the stack interacts with the protein via a van der Waals interaction with a valine residue, and not via another stacking interaction with an aromatic residue as in the other phages.

The different phages thus show a remarkable variation of how the specificity of the coat protein-RNA interaction is achieved, from several base-specific interactions in PP7 to a recognition mechanism based largely on the RNA backbone orientation in Q β . Besides the aromatic stacking that extends to the hairpin loop and a functionally conserved binding of a single adenine base in the loop, there appear to be no other common themes in the ssRNA phage coat protein-operator complexes. Still, from the currently available data, two distinct coat protein-RNA-binding modes can be recognized; the first shared by phages MS2, PRR1, and Q β and the other observed in the PP7 phage. The ssRNA phage coat protein-RNA interaction is a good example for a coevolution of protein and RNA structure, as changes in one of the components have to be complemented with corresponding changes in the other to maintain the binding. While it seems reasonable to assume that the MS2/Q β and PP7 RNA-binding modes are evolutionary related, they are very distinct, and it is difficult to envision a common ancestor and a step-by-step transition to the two RNA binding modes. It is perhaps worth mentioning that the coat protein-replicase operator interaction is not critical for the phage, as mutants with a

nonfunctional operator are still viable and only marginally less fit than the wild type (Peabody 1997; Licis et al. 2000). In addition, despite multiple experimental attempts, there is currently no evidence suggesting that an analogous interaction exists in the more distantly related phages AP205 and Cb5. Therefore, it appears that there might not be a very high pressure for the phage to conserve the interaction, and possibly, the coat protein-mediated replicase repression exists in only a subgroup of the ssRNA phages. It also cannot be excluded that the interaction has arisen more than once in different phage lineages, which might be the case for the MS2/Q β and PP7 RNA recognition mechanisms.

Interactions with the Genome in Virus Particle

The main function of the coat protein, the formation of a protein shell around the genome, also involves RNA binding. While the replicase operator hairpin is apparently the highest-affinity binding site for the coat protein, the protein is able to bind many different RNA stem-loops with lower affinity, which has been extensively characterized both biochemically and structurally. It was therefore obvious to suspect that besides the replicase operator, a number of other genomic RNA structures bind to the coat protein inside the capsid, which a protein-RNA cross-linking study using MS2 virions confirmed experimentally. The study found more than 50 potential coat protein-binding sites in the genome, most of which were clearly predicted to form a hairpin structure (Rolfsson et al. 2016). In a subsequent technological breakthrough, a medium-resolution asymmetric cryo-EM reconstruction of the MS2 bacteriophage allowed for the first time to directly visualize the genome inside the virion (Koning et al. 2016). The structure confirmed that the genome adopts a unique three-dimensional structure in the virus particles and allowed individual interactions between parts of the genome and the virion proteins to be identified. In total, 44 RNA hairpins and 33 double-stranded RNA regions were resolved that were in contact with the coat protein, while only 9 dimers did not have a nearby RNA density. Later, a higher resolution 3D reconstruction of the MS2 virion followed that identified more than 50 RNA hairpins in contact with coat protein, most of which contacted the dimers at the loop region (Dai et al. 2017). Fifteen of the coat protein-interacting stem-loops could be modeled at atomic resolution and turned out to be rather different in sequence and structure, directly demonstrating the flexibility of the coat protein in binding different RNA structures. Most of the coat protein-binding hairpins in the genomic RNA turned out to be asymmetrically distributed and predominantly located in the vicinity of the maturation protein, including the high-affinity replicase operator and two adjacent RNA stem-loops. The multiple interactions between the coat protein and hairpin structures in the genome thus lead to a model where they serve as packaging signals that together with the maturation protein help to recognize the genome and form a nucleation center for virion assembly, discussed in more detail in the next section.

Practical Applications

Besides a purely scientific interest, the ssRNA phage coat proteins and their RNA-binding properties have found a number of different applications, a significant proportion of which make use of ssRNA phage virus-like particles (VLPs). When a cloned coat protein gene is expressed in bacteria, the protein, without the need for the maturation protein or the genome, assembles into shells morphologically very similar to phage capsids (Kastelein et al. 1983; Kozlovskaya et al. 1986; Peabody 1990; Kozlovska et al. 1993). The resulting VLPs package significant amounts of cellular RNA in a largely nonspecific manner, probably via binding to different stem-loops in bacterial ribosomal and messenger RNAs (Pickett and Peabody 1993). A major area where the ssRNA phage VLPs are being explored is their use as antigen carriers in vaccine development (see Pumpens et al. 2016 and Jennings and Bachmann 2008 for reviews), and the RNA contained in the particles may activate toll-like receptors TLR3 and TLR7 that result in an enhanced immune response. The VLPs can also be obtained *in vitro* by mixing purified coat protein dimers and any heterologous RNA (Hohn 1969), which allows to package specific RNA molecules into the particles. The capsid formation can also be triggered using DNA instead of RNA, which allows to package short sequences of interest such as CpG-containing oligonucleotides into the VLPs to raise a TLR9-enhanced immune response (Bachmann et al. 2003).

A somewhat related application for ssRNA phage VLPs involves the generation of peptide display libraries using a modified version of the MS2 coat protein (Peabody et al. 2008). The method makes use of a surface-exposed region of the MS2 coat protein called the AB loop, which can tolerate short amino acid insertions without compromising its ability to assemble into VLPs. In a specifically designed vector, a randomized oligonucleotide library is ligated in-frame the AB loop, resulting in many bacterial clones each producing VLPs with a different peptide exposed on their surface. Crucially, upon assembly, the VLPs always package some coat protein mRNA into the particles, which is abundant in the cell due to the vector-driven overexpression. After affinity selection, the RNA contents of the target-bound VLPs are extracted, coat protein mRNAs amplified using reverse transcription PCR, and the corresponding peptide sequences recovered using DNA sequencing.

The high-affinity coat protein-operator hairpin interaction has been further employed to produce “armored” RNAs. Various diagnostic assays and other applications often require specific RNA sequences as controls, but due to the ubiquitous presence of ribonucleases in the environment, it is notoriously hard to avoid RNA degradation when working with naked RNA. In the armored RNA technology, the RNA of interest is engineered to contain an MS2 operator hairpin which is then produced in bacteria together with the MS2 coat protein (Pasloske et al. 1998). Due to the high specificity of the interaction, the VLPs that are assembled contain a high proportion of the operator-tagged RNA molecules inside the particles. Once the particles are assembled, the RNA is sealed from the surrounding environment, and the VLPs can then be easily purified using standard protein purification methods

without worrying about RNA degradation. The MS2 VLPs are very stable and can be stored for prolonged periods of time without special precautions. The armored RNA technology has been commercialized by Asuragen and is reviewed in detail in Mikel et al. 2015.

The specific coat protein RNA binding is also used as a research tool in molecular biology to identify or track RNA-protein interactions (see Jazurek et al. 2016 for a review). The MS2-BioTRAP method is used for identifying RNA-binding proteins (Bardwell and Wickens 1990; Tsai et al. 2011). The RNA of interest is tagged with tandem repeats of the MS2 replicase hairpin and co-expressed with a modified MS2 coat protein harboring an HB-tag which gets biotinylated *in vivo*. As a result, the RNA gets decorated with biotinylated MS2 coat protein dimers and can be captured from a cell extract using streptavidin-coated beads. Proteins bound to the RNA of interest can then be stripped off and identified using mass spectrometry or other suitable technique. A related approach can be used to track RNA molecules of interest in living cells. The RNAs are likewise tagged with MS2 operator stem-loops, but the MS2 coat protein is fused with a fluorescent tag such as the green fluorescent protein. The RNAs can then be imaged in confocal fluorescent microscopy to follow the tagged RNAs throughout the cell (Bertrand et al. 1998). The distinct specificities of the MS2 and PP7 coat proteins also allow tracking of two different RNA molecules simultaneously using different fluorescent tags.

Maturation Protein-RNA Interactions

The maturation protein, sometimes referred to as the “A” or “A2” protein in different ssRNA phages, is the least understood of the phage proteins. The maturation protein binds to the genomic RNA and gets incorporated into the capsid along with it, where it later serves as an attachment protein that mediates the binding of the virion to bacterial pili and genome ejection and entry into the host cell. All of the known ssRNA phage maturation proteins are insoluble in an isolated form which has greatly hampered their studies, and for many decades, molecular details explaining how the maturation protein accomplishes any of its different functions had remained unknown. However, a recent high-resolution crystal structure of the maturation protein from bacteriophage Q β (Rumnieks and Tars 2017) and medium- to high-resolution asymmetric cryo-EM reconstructions of whole MS2 (Dai et al. 2017) and Q β (Gorzelnik et al. 2016) virions are finally starting to provide some answers about how the ssRNA phage maturation proteins look like and function.

The structural studies have revealed that the ssRNA phage maturation proteins have a rather peculiar highly elongated and bent shape and incorporate into the virion by taking place of a single coat protein dimer in the otherwise symmetrical protein shell (Fig. 13.2a). The maturation protein has a roughly globular α -helical part that faces the capsid interior and an elongated, relatively flat β -part that interacts with the coat protein and points away from the particle at a shallow angle. In phages MS2 and Q β , both the coat and maturation proteins have approximately 20%

sequence identity, but while the coat protein structure in the two phages is very similar, the same does not hold true for the maturation proteins. Among the two proteins, only the core four-helix bundle of the α -helical region is clearly conserved, while the differences in other parts of the proteins are often too large for a reliable structural alignment. The structure of more distantly ssRNA phage maturation proteins is presumably even more distinct, although all of them probably have an inward-facing α -helical part and a surface-exposed β -region, as suggested by secondary structure predictions.

Like the other ssRNA phage proteins, also the maturation protein is a specific RNA-binding protein, and the high-resolution cryo-EM structure of the MS2 virion provides the first detailed look for any ssRNA phage at how the maturation protein binds to the RNA (Dai et al. 2017). The interaction between the MS2 maturation protein and the genome is rather extensive and involves two distinct RNA-binding surfaces on the protein and four regions in the genome (Fig. 13.7a–c). The first RNA-binding surface is located toward the distal part of the α -helical region and makes contact with two double-helical hairpin stems in the replicase coding region. The interaction is sequence-nonspecific and involves electrostatic interactions between the sugar-phosphate backbone and positively charged residues. The other RNA-binding surface is located around the central part of the protein and is composed of a portion of the central β -sheet and an adjacent part of the α -helical region. It binds two hairpins in the 3'-untranslated region of the genome, the first of which is very short and poorly defined in the structure, while the other is the very 3'-terminal hairpin where the interaction is clearly sequence-specific. Interestingly, binding of the 3'-terminal hairpin is accomplished cooperatively between the maturation protein and two adjacent coat protein dimers (Fig. 13.7d), and a small part of the RNA hairpin is directly exposed to the exterior of the virion (Fig. 13.7e).

The MS2 virion structure also provides important clues about the possible assembly pathway of the virus particle. The loop of the 3'-terminal helix contains a sequence CUGCUU that is fairly conserved among different RNA phages. In the MS2 virion, this sequence forms the stretch that is bound cooperatively by the maturation protein and two adjacent coat protein dimers and is partly exposed on the virion surface. In the Q β genome, the nucleotide stretch in the 3'-terminal hairpin loop has been shown to form a pseudoknot with a sequence within the replicase coding region, the disruption of which abolishes replication (Klovins and van Duin 1999). The likely explanation for this is that the pseudoknot together with another nearby long-distance interaction positions the M site in an orientation that allows the replicase complex to bind to the genome and initiate replication. It is not clear whether an equivalent pseudoknot is formed in the MS2 genome, but binding of the maturation protein to the 3'-terminal hairpin clearly renders the nearby 3'-terminus of the genome inaccessible to the replicase and prevents its replication. The cryo-EM structures also revealed that a higher proportion of the stem-loops that bind to the coat protein are located close to the maturation protein compared to the opposite side of the particle. In particular, three adjacent RNA stem-loops were resolved in

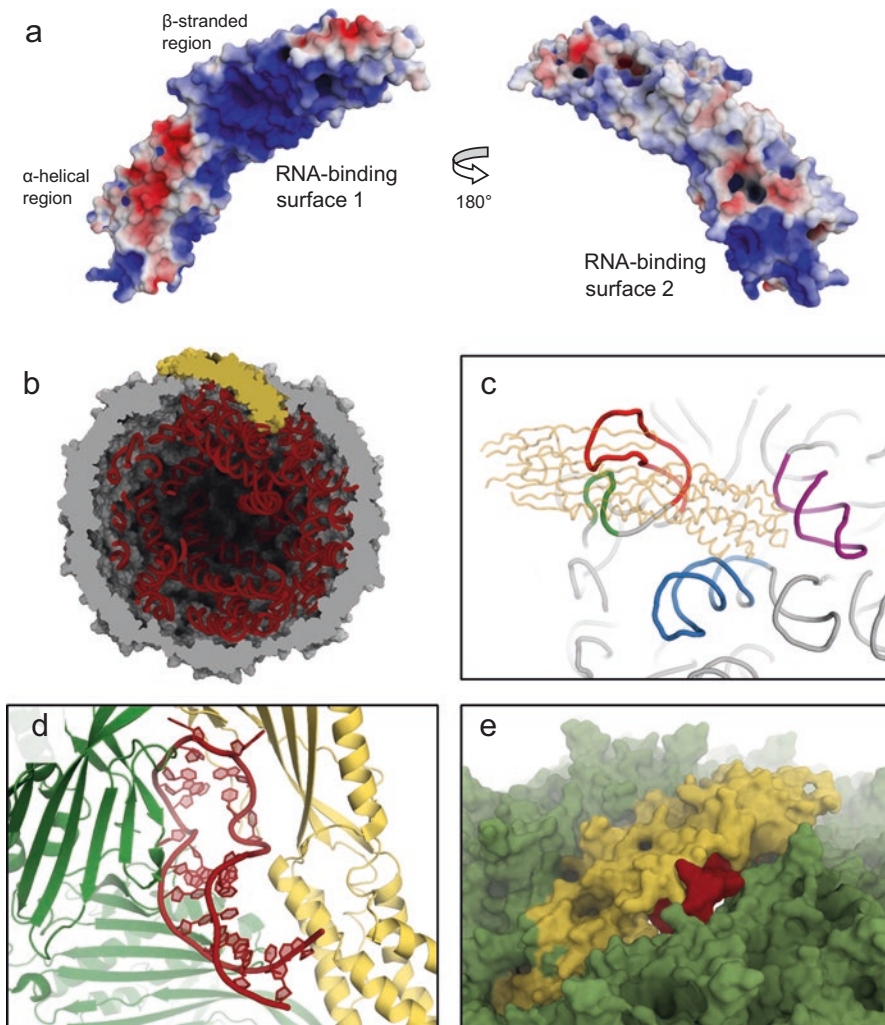


Fig. 13.7 Interactions between the maturation protein and the genomic RNA. **(a)** The MS2 maturation protein has a banana-like shape and consists of an α -helical and a β -stranded region. On the electrostatic surface of the protein, two distinct positively charged areas are present that form the two RNA-binding regions of the protein. **(b)** A cut-away view of the MS2 virion. The maturation protein (light orange in all panels) is partially exposed on the virion surface and makes contact both with the coat protein (gray) and the genome (red). **(c)** Parts of the MS2 genome in contact with the maturation protein. The protein binds to four distinct RNA stem-loops in the genome (colored). Of these, only binding to the 3'-terminal hairpin (red) is sequence-specific. **(d)** A close-up view of the interaction with the 3'-terminal hairpin. The hairpin (red) is bound cooperatively by the maturation protein and two neighboring coat protein dimers (green). In the crevice between the maturation and coat proteins, the RNA is partly exposed onto the virion surface **(e)**

the MS2 genome that bind to coat protein dimers nearby the maturation protein, among those the high-affinity replicase operator. Together, these observations suggest that binding of the maturation protein and two coat protein dimers to the 3'-terminal hairpin likely forms the nucleation center that marks the genome for packaging in new virus particles, and nearby high-affinity stem-loops then recruit more coat protein dimers that result in a rapid formation of a protein shell around the genome.

Together with the high-resolution crystal structure of the Q β maturation protein, the currently available medium-resolution cryo-EM structure of the Q β virion (Gorzelnik et al. 2016) allows to explore some of the maturation protein-RNA interaction also in this phage. Like MS2, also the Q β maturation protein has two distinctive RNA-binding regions located similarly as in the MS2 protein (Rumnieks and Tars 2017). The α -helical region likewise binds several double-helical features in the genome, and the central RNA-binding surface interacts with a single long RNA hairpin, but due to the limited resolution, it is not possible to tell if it this is also the 3'-terminal one like in MS2. The long hairpin approaches the Q β maturation protein from a different direction, and while it also appears to make contacts with a nearby coat protein dimer, these are not nearly as extensive as in MS2. In Q β , also no part of the genome becomes exposed to the outer surface of the particle, as the corresponding gap in the virion is plugged by the N-terminal part of the maturation protein that is folded differently than in MS2. Thus, while the non-atomic resolution of the Q β virion map certainly leaves a room for interpretation errors, the RNA binding of the Q β maturation appears to be sufficiently different from that of MS2, again suggesting that there might be little conservation and the RNA binding is likely even more distinctive for increasingly further related phages.

Concluding Remarks

It has become increasingly clear that the genetic material of an organism, or a virus, cannot be considered merely as the storage medium for encoding its proteins, and rarely it is more evident than in the case of the ssRNA bacteriophages. Here, the genome is the most central figure that via its complex and dynamic three-dimensional shape orchestrates the phage and host proteins to get replicated and propagated. The story of the phage life cycle is, for the most part, a story of RNA structure and different protein-RNA interactions.

The recent asymmetric cryo-EM reconstructions of ssRNA phage virions have been a major breakthrough that has allowed for the first time to directly visualize the genome inside their particles, revealing its complex and folded three-dimensional structure. We are now closer than ever to a truly molecular-level understanding of the small RNA phages, but there is certainly still much substance for further studies. Among other things, there are some blank spots left of how the ssRNA phage replicases recognize the genomic RNA, and virtually nothing is known about the molecular mechanism of how the maturation protein guides the genome into the

host cell. Furthermore, due to the advances in modern sequencing technologies, the true ssRNA phage diversity in nature is just being discovered, which stretches far beyond the few model phages on which almost all of our current understanding of these viruses is built. Surely, many more protein-RNA interactions await their discovery and exploration.

References

- Avota E, Berzins V, Grens E, Vishnevsky Y, Luce R, Biebricher CK (1998) The natural 6 S RNA found in Qbeta-infected cells is derived from host and phage RNA. *J Mol Biol* 276(1):7–17
- Bachmann MF, Storni T, Maurer P, Tissot A, Schwarz K, Meijerink E, Lipowsky G, Pumpens P, Cielens I and Renhofs R (2003) Packaging of immunostimulatory cpG into virus-like particles: method of preparation and use. E. P. Office. WO03024481 (A2)
- Bardwell VJ, Wickens M (1990) Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res* 18(22):6587–6594
- Bernhardt TG, Wang IN, Struck DK, Young R (2001) A protein antibiotic in the phage Qbeta virion: diversity in lysis targets. *Science* 292(5525):2326–2329
- Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM (1998) Localization of ASH1 mRNA particles in living yeast. *Mol Cell* 2(4):437–445
- Biebricher CK, Luce R (1993) Sequence analysis of RNA species synthesized by Qbeta replicase without template. *Biochemistry* 32(18):4848–4854
- Biebricher CK, Eigen M, Luce R (1986) Template-free RNA synthesis by Qb replicase. *Nature* 321:89–91
- Blumenthal T, Landers TA, Weber K (1972) Bacteriophage Qb replicase contains the protein biosynthesis elongation factors EF Tu and EF Ts. *Proc Natl Acad Sci U S A* 69(5):1313–1317
- Boni IV, Artamonova VS, Dreyfus M (2000) The last RNA-binding repeat of the Escherichia coli ribosomal protein S1 is specifically involved in autogenous control. *J Bacteriol* 182(20):5872–5879
- Bradley DE (1966) The structure and infective process of a Pseudomonas Aeruginosa bacteriophage containing ribonucleic acid. *J Gen Microbiol* 45:83–96
- Chao JA, Patskovsky Y, Almo SC, Singer RH (2008) Structural basis for the coevolution of a viral RNA-protein complex. *Nat Struct Mol Biol* 15(1):103–105
- Chetverin AB, Spirin AS (1995) Replicable RNA vectors: prospects for cell-free gene amplification, expression, and cloning. *Prog Nucleic Acid Res Mol Biol* 51:225–270
- Chetverin AB, Chetverina HV, Munishkin AV (1991) On the nature of spontaneous RNA synthesis by Qbeta replicase. *J Mol Biol* 222(1):3–9
- Chetverina HV, Chetverin AB (1993) Cloning of RNA molecules in vitro. *Nucleic Acids Res* 21(10):2349–2353
- Crawford EM, Gesteland RF (1964) The adsorption of bacteriophage R17. *Virology* 22:165–167
- Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, Sun R (2017) In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* 541(7635):112–116
- Dent KC, Thompson R, Barker AM, Hiscox JA, Barr JN, Stockley PG, Ranson NA (2013) The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release. *Structure* 21(7):1225–1234
- Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Vandenberghe A, Volckaert G, Ysebaert M (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* 260:500–507
- Franze de Fernandez MT, Eoyang L, August JT (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature* 219(5154):588–590

- Franze de Fernandez MT, Hayward WS, August JT (1972) Bacterial proteins required for replication of phage Q β ribonucleic acid. Purification and properties of host factor I, a ribonucleic acid-binding protein. *J Biol Chem* 247(3):824–831
- Goessens WHF, Driessen AJM, Wilschut J, van Duin J (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
- Gozdzinski KV, Cui Z, Reed CA, Jakana J, Young R, Zhang J (2016) Asymmetric cryo-EM structure of the canonical Allovivivirus Q β reveals a single maturation protein and the genomic ssRNA in situ. *Proc Natl Acad Sci U S A* 113(41):11519–11524
- Grahn E, Stonehouse NJ, Adams CJ, Fridborg K, Beigelman L, Matulic-Adamic J, Warriner SL, Stockley PG, Liljas L (2000) Deletion of a single hydrogen bonding atom from the MS2 RNA operator leads to dramatic rearrangements at the RNA-coat protein interface. *Nucleic Acids Res* 28:4611–4616
- Grahn E, Moss T, Helgstrand C, Fridborg K, Sundaram M, Tars K, Lago H, Stonehouse NJ, Davis DR, Stockley PG, Liljas L (2001) Structural basis of pyrimidine specificity in the MS2 RNA hairpin-coat-protein complex. *RNA* 7(11):1616–1627
- Gralla J, Steitz JA, Crothers DM (1974) Direct physical evidence for secondary structure in an isolated fragment of R17 bacteriophage mRNA. *Nature* 248:204–208
- Haruna I, Spiegelman S (1965) Specific template requirements of RNA replicases. *Proc Natl Acad Sci U S A* 54(2):579–587
- Hecker N, Wiegels T, Torda AE (2013) RNA secondary structure diagrams for very large molecules: RNAfold. *Bioinformatics* 29(22):2941–2942
- Helgstrand C, Grahn E, Moss T, Stonehouse NJ, Tars K, Stockley PG, Liljas L (2002) Investigating the structural basis of purine specificity in the structures of MS2 coat protein RNA translational operator hairpins. *Nucleic Acids Res* 30(12):2678–2685
- Hofstetter H, Monstein H, Weissmann C (1974) The readthrough protein A1 is essential for the formation of viable Q β particles. *Biochim Biophys Acta* 374:238–251
- Hohn T (1969) Role of RNA in the assembly process of bacteriophage ϕ . *J Mol Biol* 43:191–200
- Jazurek M, Ciesiolka A, Starega-Roslan J, Bilinska K, Krzyzosiak WJ (2016) Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases. *Nucleic Acids Res* 44(19):9050–9070
- Jennings GT, Bachmann MF (2008) The coming of age of virus-like particle vaccines. *Biol Chem* 389(5):521–536
- Kacian DL, Mills DR, Kramer FR, Spiegelman S (1972) A replicating RNA molecule suitable for a detailed analysis of extracellular evolution and replication. *Proc Natl Acad Sci U S A* 69(10):3038–3042
- Kamen R, Kondo M, Romer W, Weissmann C (1972) Reconstitution of Q β replicase lacking subunit with protein-synthesis-interference factor i. *Eur J Biochem* 31(1):44–51
- Karnik S, Billetter M (1983) The lysis function of RNA bacteriophage Q β is mediated by the maturation (A2) protein. *EMBO J* 2:1521–1526
- Kastelein RA, Berkhout B, Overbeek GP, van Duin J (1983) Effect of the sequences upstream from the ribosome-binding site on the yield of protein from the cloned gene for phage MS2 coat protein. *Gene* 23:245–254
- Kazaks A, Voronkova T, Rūmnieks J, Dishlers A, Tars K (2011) Genome structure of caulobacter phage ϕ Cb5. *J Virol* 85(9):4628–4631
- Kidmose RT, Vasiliev NN, Chetverin AB, Andersen GR, Knudsen CR (2010) Structure of the Q β replicase, an RNA-dependent RNA polymerase consisting of viral and host proteins. *Proc Natl Acad Sci U S A* 107(24):10884–10889
- Klovins J, van Duin J (1999) A long-range pseudoknot in Q β RNA is essential for replication. *J Mol Biol* 294(4):875–884
- Klovins J, Berzins V, van Duin J (1998) A long-range interaction in Q β RNA that bridges the thousand nucleotides between the M-site and the 3' end is required for replication. *RNA* 4(8):948–957

- Klovins J, Overbeek GP, van den Worm SH, Ackermann HW, van Duin J (2002) Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. *J Gen Virol* 83(Pt 6):1523–1533
- Koning RI, Gomez-Blanco J, Akopjana I, Vargas J, Kazaks A, Tars K, Carazo JM, Koster AJ (2016) Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nat Commun* 7:12524
- Kozlovskaya TM, Cielens I, Dreilinn D, Dislers A, Baumanis V, Ose V, Pumpens P (1993) Recombinant RNA phage Qbeta capsid particles synthesized and self-assembled in *Escherichia coli*. *Gene* 137(1):133–137
- Kozlovskaya TM, Pumpen PP, Dreilina DE, Tsimanis AJ, Ose VP, Tsibinogin VV, Gren EJ (1986) Formation of capsid-like structures as a result of expression of coat protein gene of RNA phage fr. *Dokl Akad Nauk SSSR* 287:452–455
- Krishnamurthy SR, Janowski AB, Zhao G, Barouch D, Wang D (2016) Hyperexpansion of RNA bacteriophage diversity. *PLoS Biol* 14(3):e1002409
- Li D, Wei T, Abbott CM, Harrich D (2013) The unexpected roles of eukaryotic translation elongation factors in RNA virus replication and pathogenesis. *Microbiol Mol Biol Rev* 77(2):253–266
- Licis N, Balklava Z, van Duin J (2000) Forced retroevolution of an RNA bacteriophage. *Virology* 271:298–306
- Lim F, Peabody DS (2002) RNA recognition site of PP7 coat protein. *Nucleic Acids Res* 30(19):4138–4144
- Lim F, Spingola M, Peabody DS (1996) The RNA-binding site of bacteriophage Qb coat protein. *J Biol Chem* 271:31839–31845
- Lowary PT, Uhlenbeck OC (1987) An RNA mutation that increases the affinity of an RNA-protein interaction. *Nucleic Acids Res* 15:10483–10493
- Meyer F, Weber H, Weissmann C (1981) Interaction of Qb replicase with Qb RNA. *J Mol Biol* 153:631–660
- Mikel P, Vasickova P, Kralik P (2015) Methods for preparation of MS2 phage-like particles and their utilization as process control viruses in RT-PCR and qRT-PCR detection of RNA viruses from food matrices and clinical specimens. *Food Environ Virol* 7:96
- Miranda G, Schuppli D, Barrera I, Hausherr C, Sogo JM, Weber H (1997) Recognition of bacteriophage Qbeta plus strand RNA as a template by Qbeta replicase: role of RNA interactions mediated by ribosomal proteins S1 and host factor. *J Mol Biol* 267(5):1089–1103
- Moody MD, Burg JL, DiFrancesco R, Lovern D, Stanick W, Lin-Goerke J, Mahdavi K, Wu Y, Farrell MP (1994) Evolution of host cell RNA into efficient template RNA by Qbeta replicase: the origin of RNA in untemplated reactions. *Biochemistry* 33(46):13836–13847
- Munishkin AV, Voronin LA, Chetverin AB (1988) An in vivo recombinant RNA capable of autocatalytic synthesis by Qbeta replicase. *Nature* 333(6172):473–475
- Munishkin AV, Voronin LA, Ugarov VI, Bondareva LA, Chetverina HV, Chetverin AB (1991) Efficient templates for Q beta replicase are formed by recombination from heterologous sequences. *J Mol Biol* 221(2):463–472
- Pasloske BL, Walkerpeach CR, Obermoeller RD, Winkler M, DuBois DB (1998) Armored RNA technology for production of ribonuclease-resistant viral RNA controls and standards. *J Clin Microbiol* 36:3590–3594
- Peabody DS (1990) Translational repression by bacteriophage MS2 coat protein expressed from a plasmid. A system for genetic analysis of a protein-RNA interaction. *J Biol Chem* 265:5684–5689
- Peabody DS (1997) Role of the coat protein-RNA interaction in the life cycle of bacteriophage MS2. *Mol Gen Genet* 254:358–364
- Peabody DS, Manifold-Wheeler B, Medford A, Jordan SK, do Carmo Caldeira J, Chackerian B (2008) Immunogenic display of diverse peptides on virus-like particles of RNA phage MS2. *J Mol Biol* 380(1):252–263
- Persson M, Tars K, Liljas L (2013) PRR1 coat protein binding to its RNA translational operator. *Acta Crystallogr D Biol Crystallogr* 69(Pt 3):367–372

- Pickett GG, Peabody DS (1993) Encapsulation of heterologous RNAs by bacteriophage MS2 coat protein. *Nucleic Acids Res* 21:4621–4626
- Pumpens P, Renhofa R, Dishlers A, Kozlovska T, Ose V, Pushko P, Tars K, Grens E, Bachmann MF (2016) The true story and advantages of RNA phage capsids as Nanotools. *Intervirology* 59(2):74–110
- Rolfsson O, Middleton S, Manfield IW, White SJ, Fan B, Vaughan R, Ranson NA, Dykeman E, Twarock R, Ford J, Kao CC, Stockley PG (2016) Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. *J Mol Biol* 428(2 Pt B):431–448
- Rumnieks J, Tars K (2012) Diversity of pili-specific bacteriophages: genome sequence of IncM plasmid-dependent RNA phage M. *BMC Microbiol* 12:277
- Rumnieks J, Tars K (2014) Crystal structure of the bacteriophage qbeta coat protein in complex with the RNA operator of the replicase gene. *J Mol Biol* 426(5):1039–1049
- Rumnieks J, Tars K (2017) Crystal structure of the maturation protein from bacteriophage Qbeta. *J Mol Biol* 429(5):688–696
- Schmidt JM, Stanier RY (1965) Isolation and characterization of bacteriophages active against stalked bacteria. *J Gen Microbiol* 39:95–107
- Schuppli D, Miranda G, Tsui HC, Winkler ME, Sogo JM, Weber H (1997) Altered 3'-terminal RNA structure in phage Qbeta adapted to host factor-less *Escherichia coli*. *Proc Natl Acad Sci U S A* 94(19):10239–10242
- Schuppli D, Miranda G, Qiu S, Weber H (1998) A branched stem-loop structure in the M-site of bacteriophage Qbeta RNA is important for template recognition by Qbeta replicase holoenzyme. *J Mol Biol* 283(3):585–593
- Schuppli D, Georgijevic J, Weber H (2000) Synergism of mutations in bacteriophage Qbeta RNA affecting host factor dependence of Qbeta replicase. *J Mol Biol* 295(2):149–154
- Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ (2016) Redefining the invertebrate RNA virosphere. *Nature* 540:539
- Shiba T, Miyake T (1975) New type of infectious complex of *E. coli* RNA phage. *Nature* 254:157–158
- Skripkin EA, Adhin MR, de Smit MH, van Duin J (1990) Secondary structure of bacteriophage MS2. Conservation and biological significance. *J Mol Biol* 211:447–463
- Sumper M, Luce R (1975) Evidence for de novo production of self-replicating and environmentally adapted RNA structures by bacteriophage Qbeta replicase. *Proc Natl Acad Sci U S A* 72(1):162–166
- Takeshita D, Tomita K (2010) Assembly of Q{beta} viral RNA polymerase with host translational elongation factors EF-Tu and -Ts. *Proc Natl Acad Sci U S A* 107(36):15733–15738
- Takeshita D, Tomita K (2012) Molecular basis for RNA polymerization by Qbeta replicase. *Nat Struct Mol Biol* 19(2):229–237
- Takeshita D, Yamashita S, Tomita K (2014) Molecular insights into replication initiation by Qbeta replicase using ribosomal protein S1. *Nucleic Acids Res* 42(16):10809–10822
- Tsai BP, Wang X, Huang L, Waterman ML (2011) Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. *Mol Cell Proteomics* 10(4):M110 007385
- Valegård K, Liljas L, Fridborg K, Unge T (1990) The three-dimensional structure of the bacterial virus MS2. *Nature* 345:36–41
- Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* 371:623–626
- Valegård K, Murray JB, Stonehouse NJ, van den Worm S, Stockley PG, Liljas L (1997) The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol* 270:724–738
- Van Duin J, Tsareva N (2006) Single-stranded RNA phages. In: Calendar R (ed) *The bacteriophages*. Oxford University Press, New York, pp 175–196

- Vasilyev NN, Kutlubaeva ZS, Ugarov VI, Chetverina HV, Chetverin AB (2013) Ribosomal protein S1 functions as a termination factor in RNA synthesis by Qbeta phage replicase. *Nat Commun* 4:1781
- Wahba AJ, Miller MJ, Niveleau A, Landers TA, Carmichael GG, Weber K, Hawley DA, Slobin LI (1974) Subunit I of Qbeta replicase and 30 S ribosomal protein S1 of *Escherichia coli*. Evidence for the identity of the two proteins. *J Biol Chem* 249(10):3314–3316
- Weber H (1976) The binding site for coat protein on bacteriophage Qb RNA. *Biochim Biophys Acta* 418:175–183
- Weber H, Weissmann C (1970) The 3'-termini of bacteriophage Qbeta plus and minus strands. *J Mol Biol* 51(2):215–224
- Weiner AM, Weber K (1971) Natural read-through at the UGA termination signal of Qbeta coat protein cistron. *Nat New Biol* 234:206–209
- Weissmann C, Feix G, Slor H, Pollet R (1967) Replication of viral RNA. XIV. Single-stranded minus strands as template for the synthesis of viral plus strands in vitro. *Proc Natl Acad Sci U S A* 57(6):1870–1877
- Winter RB, Gold L (1983) Overproduction of bacteriophage Qb maturation (A2) protein leads to cell lysis. *Cell* 33:877–885
- Witherell GW, Uhlenbeck OC (1989) Specific RNA binding by Qb coat protein. *Biochemistry* 28:71–76
- Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9(1):133–148