Structural Constraints and Mutational Bias in the Evolutionary Restoration of a Severe Deletion in RNA Phage MS2

Normunds Licis,[1] **Jan van Duin**[2]

[1] Biomedical Research and Study Centre, University of Latvia, Ratsupites 1, LV 1067 Riga, Latvia
[2] Leiden Institute of Chemistry, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands

Abstract. A 4-nucleotide (nt) deletion was made in the 36-nt-long intercistronic region separating the coat and replicase genes of the single-stranded RNA phage MS2. This region is the focus of several RNA structures conferring high fitness. One such element is the operator hairpin, which, in the course of infection, will bind a coat-protein dimer, thereby precluding further replicase synthesis and initiating encapsidation. Another structure is a long-distance base pairing (MJ) controlling replicase expression. The 4-nt deletion does not directly affect the operator hairpin but it disrupts the MJ pairing. Its main effect, however, is a frame shift in the overlapping lysis gene. This gene starts in the upstream coat gene, runs through the 36-nt-long intercistronic region, and ends in the downstream replicase cistron. Here we report and interpret the spectrum of solutions that emerges when the crippled phage is evolved. Four different solutions were obtained by sequencing 40 plaques. Three had cured the frame shift in the lysis gene by inserting one nt in the loop of the operator hairpin causing its inactivation. Yet these low-fitness revertants could further improve themselves when evolved. The inactivated operator was replaced by a substitute and thereafter these revertants found several ways to restore control over the replicase gene. To allow for the evolutionary enrichment of low-probability but high-fitness revertants, we passaged lysate samples before plating. Revertants obtained in this way also restored the frame shift, but not at the expense of the operator. By taking larger and larger lysates samples for such bulk evolution, ever higher-fitness and lower-frequency revertants surfaced. Only one made it back to wild type. As a rule, however, revertants moved further and further away from the wild-type sequence because restorative mutations are, in the majority of cases, selected for their capacity to improve the phenotype by optimizing one of several potential alternative RNA foldings that emerge as a result of the initial deletion. This illustrates the role of structural constraints which limit the path of subsequent restorative mutations.

Key words: RNA phage — RNA virus — RNA structure — Translational control — Quasi-species

Introduction

RNA bacteriophage MS2 is a member of the family of Leviviridae. Virions have icosahedral symmetry and infect male *Escherichia coli* bacteria. The ssRNA genome contains 3569 nucleotides (nt) and consists of four genes (Fig. 1A). Maturation protein, present at one copy per virion, is needed for phage attachment to the F pili of male bacteria. Ninety coat-protein dimers are required to build up the capsid while lysis protein helps to break the bacterial cell wall to liberate the progeny. Replicase, finally, is the enzyme that, together with several host proteins, multiplies the RNA via the synthesis of a separate minus strand.

The amino acid sequences of these four proteins determine the sequence of the genome to a great deal. However, some freedom is provided by the degener-
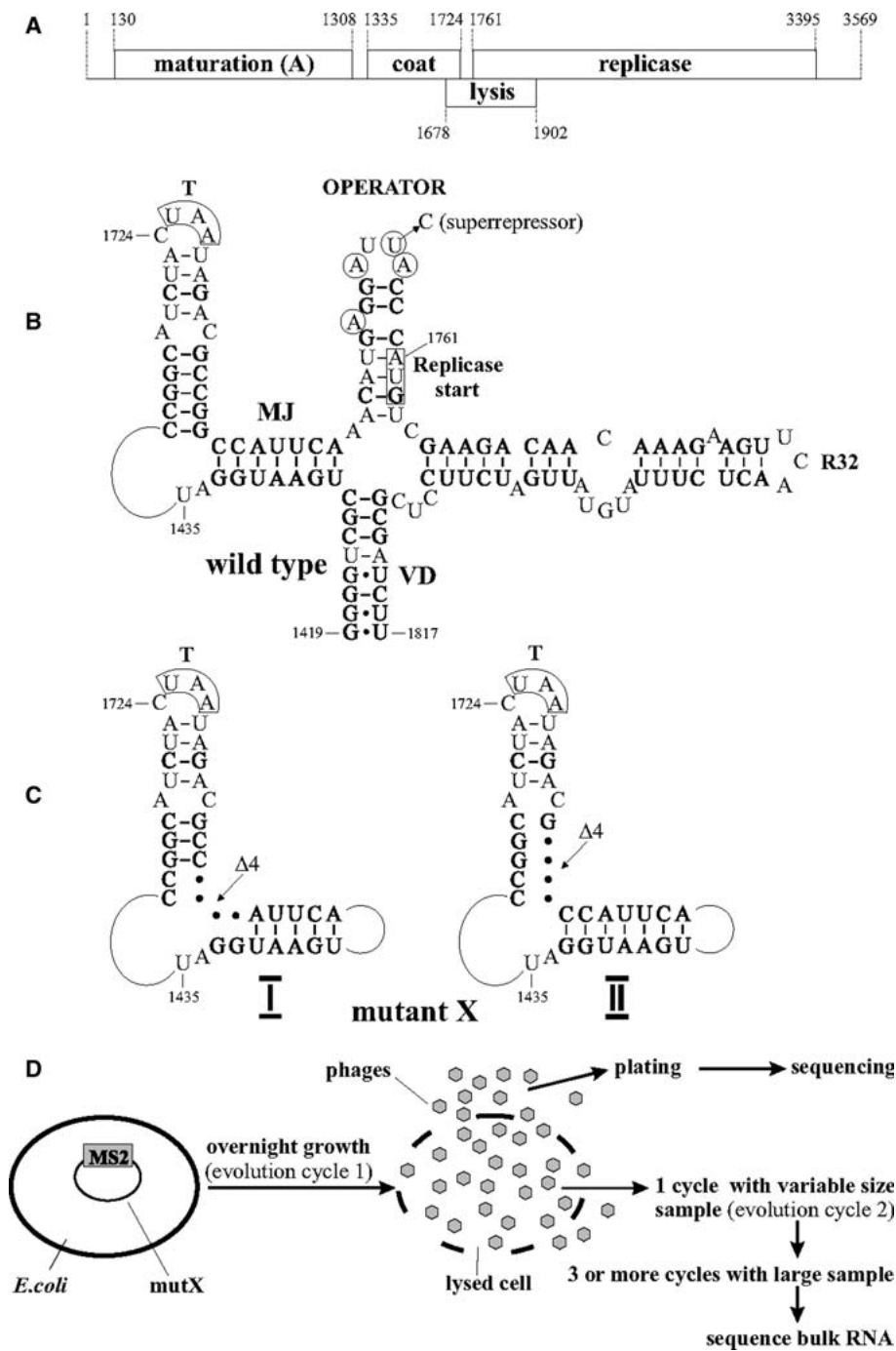
Correspondence to: N. Licis; *email:* licis@biomed.lu.lv

Fig. 1. A Genetic map of RNA phage MS2. The lysis frame is shifted +1 with respect to coat and replicase genes. **B** Relevant secondary structure between nucleotide (nt) 1419 and nt 1817. VD and MJ are long-distance interactions that make initiation of replicase synthesis dependent on translation of the coat gene. T stands for coat-terminator hairpin. The hairpin harboring the start of the replicase gene is called the operator. Encircled nt in the operator interact in a base-specific way with a coat-protein dimer. The U1756C mutation increases the affinity of the coat protein for the operator. **C** Two alternative structures that may form as a result of the deletion present in mutant X. The four dots indicate the deleted nt. Structure I is predicted to be slightly more stable than structure II. **D** Outline of the experiment.

acy of the genetic code. In addition, sequence comparison of strains shows that at multiple positions more than one amino acid is compatible with the function of the encoded protein. In more extreme cases parts of a protein are dispensable, e.g., the N-terminal half of the lysis protein may be deleted without loss of lysis function (Berkhout et al. 1985). All these sequence opportunities can be and have been used by the phage to optimize its phenotype and thus to increase its fitness under prevailing laboratory conditions. Improved fitness, meaning more progeny per infected bacterium and better timing of host lysis, is achieved by precisely adjusting the relative levels of the various gene products as well as the timing of their appearance and shutoff. In RNA phages where there is neither transcription nor production of subgenomic messengers, all of these regulations are brought about by formation and disruption of RNA secondary structure. For instance, local stem-loop structures or long-distance pairings, present at translational start sites, can, if stable enough, quench ribosomal access (de Smit and van Duin 1990a,b).

A Insertion and deletion revertants

		MJ	operator	
wt	UAAUA	GACGCCGGCCAUUC	AAACAUGAGG	AUUA CCCAUGU
mutX	UAAUA	GACGCC....AUUC	AAACAUGAGG	AUUA CCCAUGU
revC	UAAUA	GACGCC....AUUC	AAACAUGAGG	AUUA**C**CCCAUGU
revA	UAAUA	GACGCC....AUUC	AAACAUGAGG	AUUA**A**CCCAUGU
revL	UAAUA	GACGCC....AUUC	AAACAUGAGG**A**	AUUA CCCAUGU
revD	UAAUA	GACGCC....AUUC	..ACAUGAGG	AUUA CCCAUGU
revB	UAAUA	GACGCC....AUUC**A**	AAACAUGAGG	AUUA CCCAUGU
revIN1	UAAUA	GACGCC...**C**AUUC	AAACAUGAGG	AUUA CCCAUGU
revS	UAAUA**A**	GACGCC....AUUC	AAACAUGAGG	AUUA CCCAUGU

B Duplication revertants

		MJ	operator
wt	AGACGCC	GGCCAUUCAA	ACAUGAGGAUUACCCAUGU
mutX	AGACGCC	....AUUCAA	ACAUGAGGAUUACCCAUGU
revI4	AGACGCC	....AUU**CAAA**CAAACAUGAGGAUUACCCAUGU	
revIN4	AGACGCC	**AUUC**AUUCAA	ACAUGAGGAUUACCCAUGU
revIN7	**AGACGCC**	AGACGCCAUUCAA	ACAUGAGGAUUACCCAUGU
revIN10	AG**ACGCCAUUCA**ACGCCAUUCAA		ACAUGAGGAUUACCCAUGU

Fig. 2. List of the primary revertants (one change). Insertions and duplications are in boldface. Deleted nucleotides are indicated by dots. Stop and start codons of coat and replicase genes, respectively, are underlined.

RNA hairpins can also serve as binding sites for the coat and maturation protein thereby initiating and propagating encapsidation.

RNA bacteriophages are very well suited to visualize such a process of evolutionary optimization by RNA folding because their evolution proceeds very fast. This is the result of the high error rate in viral RNA replication (Drake 1993), the large number of progeny ($\sim 10^{12}$/ml culture), and the short infection cycle (several hours). Furthermore, extensive knowledge exists on how RNA structure contributes to fitness and exerts its effect on translation and coat protein binding in these viruses (de Smit 1998; Witherell et al. 1991).

One way of setting this evolutionary optimization process in motion is to destroy one or more of these regulatory RNA structures. This can be achieved, for instance, by introduction of amino acid-neutral substitutions (Olsthoorn et al. 1994; Klovins et al. 1997a), by random substitutions at RNA sections encoding nonessential amino acids (Licis et al. 1998, 2000), or by more serious lesions such as tandem nonsense codons and deletions or insertions in noncoding regions (Olsthoorn and van Duin 1996a; Klovins et al. 1997b; Licis et al. 2000).

In this report the RNA phage was disabled by a deletion of 4 nt in the intercistronic region separating coat and replicase genes. We call it mutant X (Fig. 1C) or mutX (Fig. 2A). This piece of the genome is part of the lysis gene and the 4-nt deletion leads to frame shifting. However, the absence of a lysis protein is not the only defect caused by the mutation. As we will show and explain, it also disrupts regulation of replicase synthesis. The structure presented in Fig. 1B holds the key to the control and shutoff of replicase production. Translation of the replicase gene is dependent on the readout of the coat gene. This coupling is the result of the long-distance MJ and VD base pairings (Fig. 1B). Together with the hairpins R32, the operator, and the coat terminator hairpin T, these interactions prevent vacant ribosomes to access the replicase start. That is, if coat gene translation is blocked, there will be no replicase production unless the structure encompassing the replicase initiation site is destabilized by deletions (Berkhout and van Duin 1985) or mismatches (van Himbergen et al. 1993). In a normal infection, every time the MJ and VD interactions are broken by the passage of a coat-synthesizing ribosome, the structure is sufficiently destabilized to permit binding of a vacant ribosome to the replicase start. The conditional access of ribosomes to the replicase start is an important part of the switch from translation to replication that needs to occur early in infection (for full details on this switch see van Duin [1999] and van Meerten et al. [2001]). The 4-nt deletion (Fig. 1C, left) disrupts this control mechanism by weakening the terminator and the MJ pairing; this provokes uncontrolled replicase synthesis, which leads to a reduced output of functional virus, depriving the phage of one of its high-fitness features.

Another important structure that plays a role in this study is the operator hairpin. When coat protein reaches a critical concentration a dimer specifically binds the operator, thereby denying ribosome entry and switching off any further replicase synthesis. At the same time this coat-operator complex forms the nucleation point for capsid formation. The affinity between operator and dimer, and thus the timing of encapsidation, has been precisely adjusted as witnessed by the finding of Lowary and Uhlenbeck (1987) that an operator mutant (U1756C) exists with

even higher affinity for the coat protein, indicated as "superrepressor" in Fig. 1B. Despite widely diverging sequences of operators and coat proteins, their interaction has been preserved in all RNA phages (Tars et al. 2000; Lim et al. 2001).

Because of the genetic overlap the frame-shift deletion that we have introduced can only be repaired by insertions or deletions in the region 1728 to 1761 where there is no such coding overlap.

In this paper we describe the large variety of ways in which phage MS2 can repair the damage inflicted by the deletion. The first step is always the restoration of the lysis reading frame, often through sacrificing the operator. Thereafter, evolutionary efforts are directed to repair the operator hairpin (if damaged). Finally, control of replicase gene expression is pursued. This must also be the order in which these three features contribute to fitness. We also compare the fitness of the various pseudorevertants.

Materials and Methods

Bacterial Strains and Plasmids

Escherichia coli K12 strain M5219 (M72 trpA$_{am}$, lacZ$_{am}$, Sm^r/ λdbio$_{252}$, cI857ΔH1), encoding the thermosensitive λ repressor (cI857) and the transcriptional antitermination factor N (Remaut et al. 1981), was host for plasmids. Evolution was performed in *E.coli* F$^+$ AB259 (Hfr3000, Thi$^-$, Su$^-$) cells. Bacteria were grown on LB broth.

The infectious MS2 cDNA clone contains a full-length copy of the phage genome downstream of the thermoinducible P$_L$ promoter of phage λ (Olsthoorn et al. 1994). In plasmids used for measurements of replicase gene expression the 1365–2057 MS2 cDNA fragment is located behind the P$_L$ promoter and the *lacZ* gene is fused at the *Bam*HI site (2057) to the replicase gene (van Himbergen et al. 1993). pCOAT184 is derived from pACYC184 by cloning the MS2 coat protein gene into the *Bam*HI site of the tetracycline resistance gene, and in this construct the coat protein is under control of the constitutive *tet* promoter (Berkhout 1985).

Mutant X

Mutant X was obtained accidentally. An *Xba*I-*Bfr*I (1303–1901) MS2 cDNA fragment prepared for ligation into the corresponding restriction sites of the infectious clone was also treated with *Cfr*10I recognizing the RCCGGY sequence. This restriction and the following cloning somehow resulted in deletion of the CCGG sequence.

Recording Phage Evolution

The principle of the experiment is outlined in Fig. 1D. Either the lysate obtained after transforming F$^-$ cells with the mutant X plasmid was plated to separate the genotypes for subsequent sequence and evolutionary analysis or the indicated amount of plaque forming units (pfu) was directly used as the starting material for evolution. Phage generation from the infectious cDNA in the F$^-$ host is defined as cycle 1, the first infection of F$^+$ bacteria as cycle 2, and so on. To start cycle 2, variable amounts of pfu were employed. Subsequently, about 10^5–10^6 pfu from the previous

cycles was taken to initiate cycle 3 and further, except for analyses 1.10, 1.11, 2.14, 2.15, 10.20, and 10.21 (Table 1). In these cases about 10^8–10^9 pfu was used. Most of the evolutions were carried out in duplicate. Sequence analysis of phage RNA was performed after RT-PCR amplification of the 1200–2067 MS2 cDNA fragment, routinely using a primer identical to the MS2 region 1628–1648 (Licis et al. 2000). Sequencing was carried out with the Big-Dye Termination Cycle Sequencing Ready Reaction DNA sequencing kit (Applied Biosystems), and the results were analyzed on an ABI PRISM 3100 Genetic Analyser (Applera).

Competitions

To test the relative fitness of two revertants, the titers of their corresponding lysates were determined and then mixed to give a pfu ratio of 1:1. Subsequently, the mixtures were passaged on F$^+$ bacteria for up to 10 cycles. The RNA from the mixture was then sequenced. The relative presence of each revertant was reflected by the respective band intensity on the sequence gel at the positions of sequence differences.

Measuring Replicase Gene Expression

The β-galactosidase activity of replicase/*lac*Z fusions was determined according to the standard procedure (Miller 1972) as described earlier (Licis et al. 1998). To assess the effect of the MS2 coat protein on replicase synthesis, the protein was provided *in trans* from the pCOAT184 plasmid. The results are averaged from three or four measurements.

Results

The System

Some time ago full-length cDNA of MS2 RNA was prepared and inserted in a plasmid. *E. coli* cells transformed with this plasmid produce phage spontaneously, even if the preceding P$_L$ promoter is not induced. Presumably, this results from spurious transcription of MS2 cDNA (Taniguchi et al. 1978; Olsthoorn et al. 1994). After overnight growth of such a transformed bacterial culture the supernatant contains about 10^{11} pfu/ml. To avoid reinfection these plasmid experiments are carried out with *E. coli* F$^-$ cells, which cannot be infected from without. When bacteria are transformed with MS2 cDNA mutants the titer of the supernatant is orders of magnitude lower, depending on the gravity of the mutation and on the probability that viable revertants arise by replication errors. For instance, substitutions destabilizing the MJ interaction yield a titer of about 10^{10} to 10^9 pfu/ml (Licis et al. 1998), whereas destruction of the operator hairpin by several point mutations lowers phage production to 10^8 pfu/ml (Licis et al. 2000). But even the introduction of amino acid-neutral mutations can cause a serious drop in the titer up to a factor of 10^5 (Olsthoorn et al. 1994; Klovins et al. 1997a). A titer of only 10 pfu/ml was obtained when the intercistronic region between maturation and coat genes was deleted, and here we are close to the death of the phage (Olsthoorn and van Duin 1996b).

Table 1. Revertants ranked by the sample size of the first infection

Analysis no.	Population size for first infection (evolution cycle 2); pfu	Revertants present after three additional cycles (evolution cycle 5) at large population size[a]	Revertants present after further evolution at large population size (last cycle)[a]
1.1	26	C, U > B	nd
1.2	26	C, U CC	nd
2.1	27	U, CC	nd
2.2	27	BM_3	nd
10.1	43	C, U	nd
10.2	43	D > C > B	nd
2.3	54	C > ?	nd
2.4	54	S > ?	BM_2M_3 (10)
2.5[b]	81	B	nd
2.6[b]	81	BM_2, BM_3	nd
10.3	86	B, BM_3	nd
10.4	86	B, BM_2, BM_3	nd
10.5[b]	86	B	nd
10.6	86	C	nd
10.7	86	14 > B, C	B > 14 (10)
10.8	86	C, CC ∼ B	BM_3 BM_3V > CC (20)
2.7	135	B, BM_2 ∼ C	nd
2.8	135	B ∼ C ∼ D	nd
10.9	215	B∼IN1	nd
10.10	215	B ∼ IN1 ∼ IN10	IN1 (20)
1.3	260	B, BM_2 > C	nd
1.4	260	B	nd
2.9[b]	270	B > C	nd
2.10[b]	270	IN10	nd
1.5	270	(C > B)[c] B	BV (20)
10.11	375	CC, C	nd
2.11	410	B	nd
10.12[b]	430	IN1	nd
10.13[b]	430	IN1	nd
10.14	430	IN1 ∼ BM_3	IN1 > BM_3 (7)
10.15	430	B	nd
1.6	700	B	BT_1 (20)
1.7	2600	B	nd
1.8	2600	BM_2, BM_3	nd
1.9	2700	B	BV (20)
2.12[b]	2700	IN10	nd
2.13[b]	2700	B > CC	nd
10.16	3750	B, BM_3	nd
10.17	4300	B	nd
10.17-rep	4300	IN10 > B, IN1	nd
10.18[b]	4300	B, BM_2	nd
10.19	37500	BM_2, BM_3	nd
1.10	10^5	B	nd
1.11	10^5	B ∼ IN7	IN7 (30)
2.14	10^5	B ∼ IN10 > IN1	IN10 > IN4 (30)
2.15	10^5	B > IN4	IN4C (30)
10.20	10^5	B ∼ IN1	IN4C (30)
10.21	10^5	B ∼ INI > IN4	IN4C > IN7A (30)

Note. nd, not determined. Most of the analyses were carried out in duplo, e.g., analyses 1.1 and 1.2 are two samples of the same lysate produced by clone 1. Similarly, analyses 10.3 and 10.4 are two samples of the same lysate produced by clone 10, etc. C, B, etc., stand for revC, revB, etc. (∼) Revertants present approximately equally in bulk sequence. (>) Revertant(s) dominates in the bulk sequence. (?) From the bulk sequence it is impossible to distinguish the type of minor revertant(s) present.

[a] Usually about 10^5 pfu.
[b] Experiment repeated with the same lysate sample, with the same outcome except for analyses 10.17 and 10.17-rep.
[c] After evolution cycle 2.

For all mutants described above the progeny consisted entirely of revertants. The 4-nt deletion studied here causes the titer to drop from 10^{11} to 10^5 pfu/ml. The fact that they are plaque-forming units means that all of these 10^5 phages/ml have repaired their lysis function. The ratio between the titer of

mutant X and that of wild-type (wt) MS2 cDNA reflects the probability of making a replication error that restores the lysis reading frame. As we shall see below the major repair event is the insertion of a single base at a specific place and the probability of this error is apparently 10^{-6} per genome replication. The driving force for the generation of revertants is the error-prone viral replicase. This can be concluded from our observation that mutations that interfere with the replicatability of the RNA are (close to) lethal, even if they do not change the protein sequences of the phage (Klovins et al. 1998; Klovins and van Duin 1999). The error frequency for viral replicases has been estimated at 10^{-3} to 10^{-4} per nucleotide (transitions) (Drake 1993) and it is assumed that up to 30% of the progeny may carry a mutation. Under this mutation pressure the wt sequence must apparently have a large selective advantage to be maintained in the quasi-species.

Isolation of RevC, RevA, RevL, and RevD by Plating

To reveal the spectrum of high-probability solutions found by the phage to repair its lysis frame, we sequenced 40 randomly picked plaques obtained from three cultures independently transformed with mutant X and distinguished here as clone 1, clone 2, and clone 10. Thirty-five had an extra C inserted at a row of three C's just ahead of the replicase start (revC; Fig. 2A). Three plaques, which were called revA, had an extra A residue at this same position. One plaque had an A insertion four bases further upstream but still in the loop of the operator (revL). Finally, there was one plaque, revD, that had deleted two A's from a row of three. Here, the lysis frame is restored at the expense of another amino acid from the lysis protein. There were 15 revC plaques in clone 1 and 10 revC plaques in both clone 2 and clone 10, indicating that its predominance was not the result of an early mutation occurring accidentally at this position. Apparently, slipping-back of the viral replicase at the three C's takes place relatively easily.

Improvements on RevA and RevC

RevA and revC have an extra A and C, respectively, in the loop of the operator hairpin. Although this insertion repairs the lysis frame, it inactivates the operator (see Introduction) either by forcing a loop of 5 nt or by maintaining the tetraloop but turning the bulged A at position 1751 into an A*C mismatch (Fig. 3B). This repair mutation is therefore a typical example of two steps forward, one step back (Otto 2003). The inactivation of the operator is illustrated in Fig. 4. In this figure we measure the leaky replicase synthesis that occurs in the absence of coat-gene translation. As shown, replicase translation is severely reduced in mutant X when coat protein is provided *in trans* (compare "X" with "X+coat"), producing 1084 versus 127 units of β-gal, showing the repressing effect of coat protein on replicase expression when the operator is intact. However, in revC the presence of coat protein has no repressing effect on replicase synthesis (628 versus 637 units). Clearly, the operator present in revC has lost its affinity for the coat-protein dimer. (The fact that the β-gal value of revC is about half that of mutant X most likely reflects the increased thermodynamic stability of the operator by the additional C residue turning a highly destabilizing bulge into a much less destabilizing mismatch).

The absence of a functional operator means that replicase synthesis does not come to a stop and also that encapsidation will be delayed and less efficient because the nucleation site is lost. Thus, after repair of the lysis frame, the recovery of some sort of operator becomes the next priority. We monitored the evolution of six revC and two revA plaques. In some cases revC plaques evolved identical revertants, yielding altogether four kinds of revC progeny phages. The results are summarized on a sequence scale in Fig. 3A and on a structural scale in Fig. 3B. Basically, two pathways are followed, each optimizing one of the two potential conformations of the revC and revA operator shown in Fig. 3B. The tetraloop-mismatch conformation is improved in revCC and revAC (Fig. 3B, top) by the U1756C mutation. In the wt operator this change leads to super repression because its affinity for the coat-protein dimer is even greater than in the wt (Fig. 1B). It is reasonable to assume that in these mutants it compensates, partially or fully, for the loss of the bulged A.

The other pathway optimizes the pentaloop-bulge folding (Fig. 3B, bottom). The substitutions adopted are those that turn the pentaloop into a triloop by closing the top base pair. In the A*C juxtaposition of nt 1 and nt 5 of the pentaloop of revC, either the A changed to a G to create a G-C pair (revCG) or the C at position 5 turned into a U to make an A-U pair (revU). Similarly, in revA the A*A juxtaposition in the pentaloop evolved to a U-A pair (revAU). The triloop in the revertants always contains an A in its third position. To appreciate why the pentaloop evolves to a triloop we refer to recent work of Convery et al. (1998). Here, an in vitro selection/amplification technique (SELEX) was used to evolve RNA aptamers that can bind the MS2 coat-protein dimer with a high affinity. Next to base-pair variants of the wt operator the authors obtained the F6 aptamer shown in Fig. 3B. X-ray structure analysis showed that the encircled bases have specific contacts with the coat protein. The similarity between the aptamer and
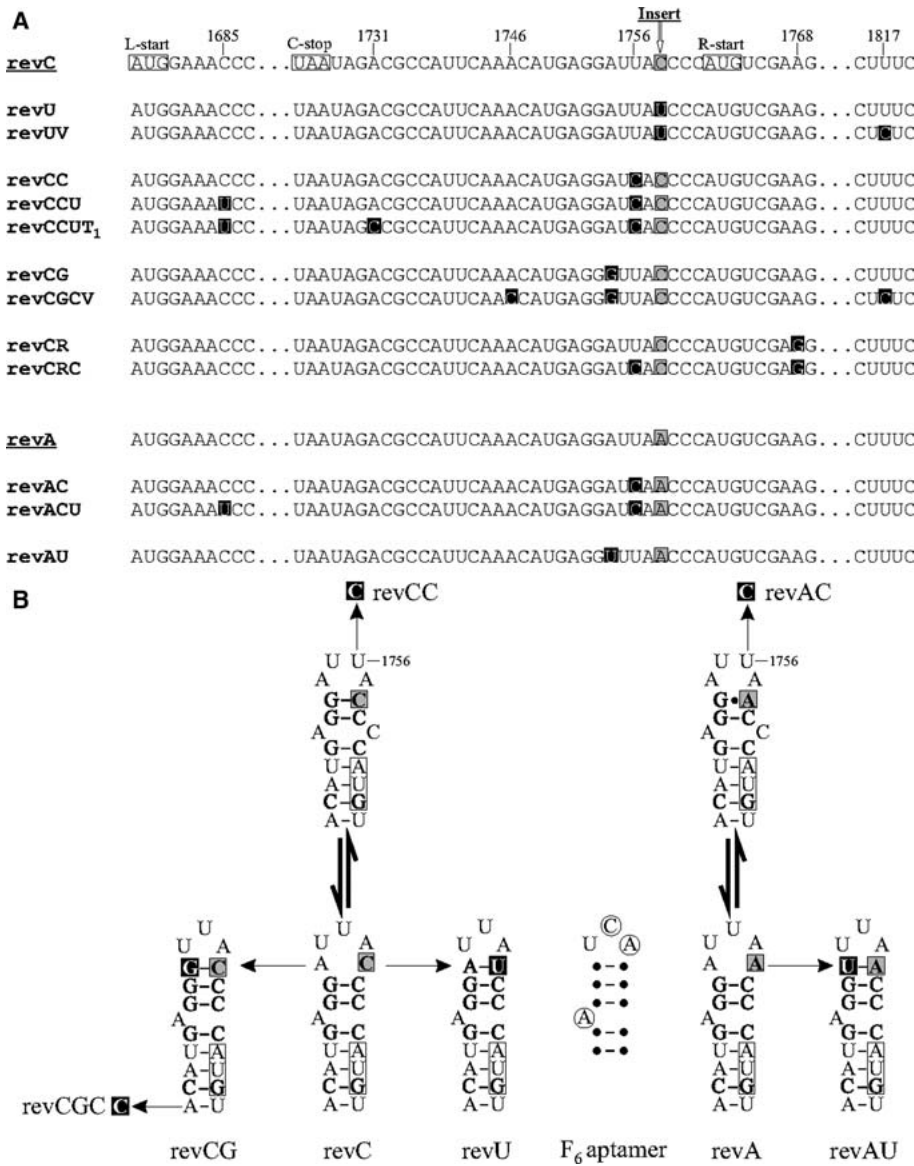
A

B

Fig. 3. **A** Evolution of revC and revA plaques. Suppressor mutations are boxed. **B** Consequences of the suppressor mutations for the operator hairpin structure. U1756C leads to a higher affinity between operator and coat protein. The structures shown as revCG, revU, and revAU are, except for the absence of a C in the loop, identical to the F6 aptamer, evolved in an in vitro selection experiment. Encircled nucleotides interact in a base-specific way with the coat protein. Note that in both operator and aptamer, any base pair will do. Revertant nomenclature is as follows: the first base in the name refers to the first mutation, e.g., revC; each subsequent mutation is added to the name, e.g., revCC, which becomes, finally, revCCU and revCCUT$_1$.

our improved operator is obvious except that we have no C as the middle base of the triloop. We may also recall our study in which almost the complete operator was randomized and then allowed to reform by evolution (Licis et al. 2000). Next to an authentic operator we obtained hairpins that turned out to be F6 aptamers, except for the second loop position, which was not always a C. Finally, we draw attention to the fact that hairpin R32 (Fig. 1B) is also an F6 aptamer and likely involved in transmitting the encapsidation process after the first dimer has docked on the true operator (van den Worm 2004).

At this point lysis frame and operator have been repaired in revC and revA. What is left to do is to restore the coat-replicase coupling. Two revertants (revUV and revCGCV) do this by the U1817C substitution (Fig. 3A). This suppressor mutation, which stabilizes the VD interaction by replacing a U*G for a C-G pair, was also the evolutionary response in a

previous study where we destabilized the MJ pairing by mismatches. Indeed, the uncoupled replicase synthesis of revUV and revCGCV is low compared to mutant X (Fig. 4) and is about that of the wt.

One other revertant, revCCUT$_1$, uses A1731C to repress replicase. This is an interesting solution. How it works will be discussed below. Then there is A1768G, found in revCR. This mutation might be explained by its potential to favor an alternative and more stable pairing in the lower part of R32 where GAGGA (1766–1770) matches UCCUC (1805–1810) (Fig. 5, center). A completely different explanation for the A1768G selection is that it changes an AAG to an AGG codon. AGG decoding tRNA is present in extremely low quantities in the *E. coli* cell and the presence of AGG codons, especially when close to the AUG start, is known to repress translation by causing queuing of ribosomes (Rosenberg et al. 1993; Gonzalez de Valdivia and Isaksson 2004).
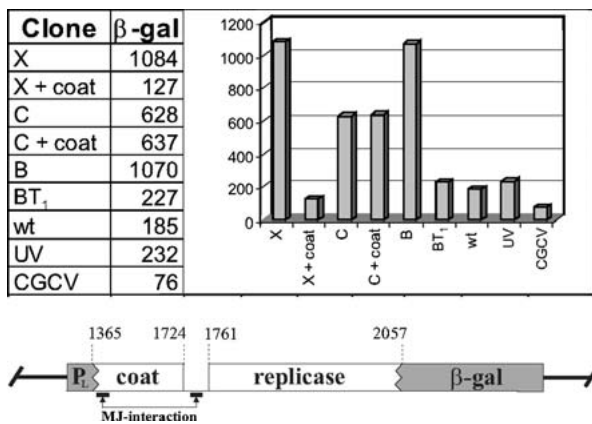
Fig. 4. Leaky translation of the replicase gene in the wild type, in mutant X, and in various revertants. The replicase is fused to the β-galactosidase gene as indicated. The start of the coat gene has been removed but the sequences for the MJ and VD interactions are still present. We therefore measure the leakage of replicase synthesis. P_L is a heat-inducible promoter. C, B, etc., stand for revB, revC, etc.

Still other revertants accumulate restorative mutations that we cannot easily interpret. First, there is C1685U. This mutation, which we found in two independent lines of descent, weakens the lysis hairpin and is predicted to enhance lysis protein synthesis. Its meaning is not clear. Finally, there is A1746C in rev-CGCV. Possibly, this mutation serves to readjust the stability of the operator which has become too strong by the three consecutive G-C pairs in the top (Fig. 3B).

Appearance of RevB by Passaging of Lysate

To uncover low-probability revertants by plating one needs to sequence very large numbers of plaques. However, one can enrich for the presence of low-frequency but high-fitness revertants by passaging bulk lysate. Accordingly, a restricted amount of lysate was allowed one cycle and then passaged for three additional cycles at a large population size (see Fig. 1D and Materials and Methods). Thereafter, bulk phage RNA was sequenced to reveal the presence of a revertant(s). As would be expected, when amounts of lysate, corresponding to only ∼25 plaques were evolved this way, revC remained the dominant revertant. However, starting with larger amounts of lysate resulted in the presence of a new solution, revB (Table 1). RevB has added an extra A to a row of three to solve the frame-shift problem (Fig. 2A). This insertion, although unlikely to occur, must be a better solution for the 4-nt deletion, as it outgrew the revertants identified by plaque analysis.

Evolution of RevB

RevB was the first low-frequency, high-fitness revertant obtained when a small amount of lysate was
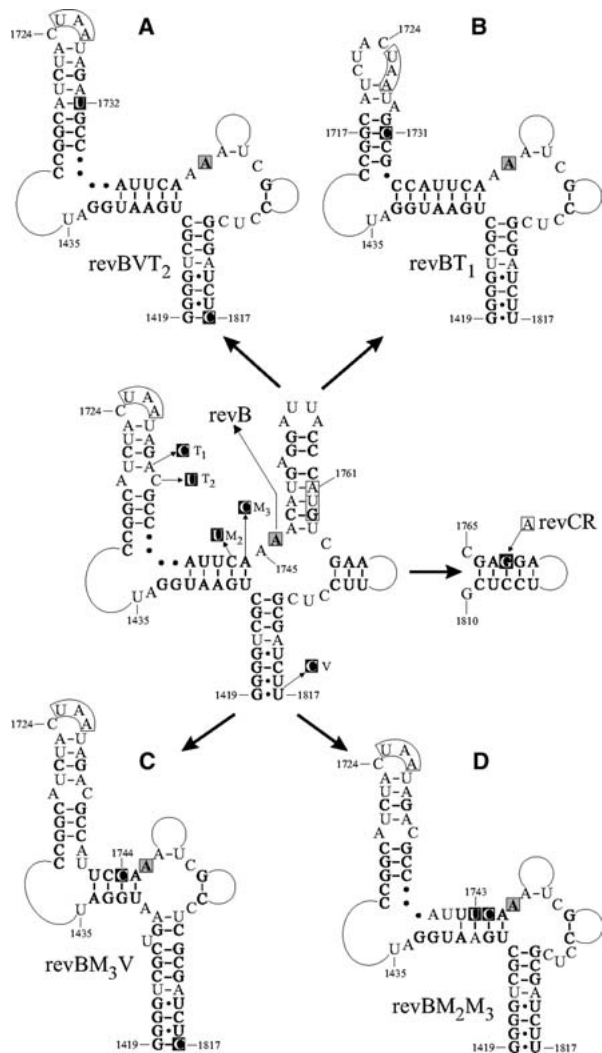
Fig. 5. Predicted RNA secondary structure in four different developments of revB. In the center we show the acquired suppressor mutations in the context of the original RNA structure. In **B**, **C**, and **D** the suppressor mutations are shown in the context of the new RNA structure that they induce. Note that the dots in panels **B** and **D** do not indicate actual deleted nucleotides (nt). They only show how many nt are missing to form a wild-type structure. RevBVT$_2$ (**A**) descended from revBV, not in the bulk evolutions presented in Table 1 but in another passaging. RevCR, shown at the right, is a descendent of revC.

passaged to enrich for such phages (see Table 1; e.g., analyses 1.1, 2.2, 10.2, etc.). RevB was clearly able to displace revC after five cycles of passaging (analysis 2.5 and further down). The molecular inferiority of revC with respect to revB is not difficult to understand; the initial insertion in revC inactivates the operator hairpin, which is a high-fitness feature. On the other hand, revB has the insertion at a seemingly neutral place just at the border of operator and MJ structure (Fig. 2A and Fig. 5, center). Still, revB suffers from a destabilized MJ pairing leading to leaky replicase production. This loss of control is illustrated in Fig. 4, where production of replicase is measured. As before, the cDNA construct used for this experi-

ment does not contain the coat-gene start and therefore the uncoupled, leaky replicase translation is determined. There are 185 units of β-gal activity (leakage) in the wt construct but it is 1084 units in mutant X. At the same time it is clear and to be expected that the extra A residue in revB does nothing to help control replicase. Indeed, revB exhibits the same level of uncontrolled replicase synthesis as mutant X (1070 units; Fig. 4).

Further evolution of revB reveals that there are several ways to regain control over replicase synthesis (Fig. 5, center). In the first one, revBV, the U1817C substitution, which we have already encountered in revUV and revCGCV, strengthens the VD interaction as it replaces G*U by G-C (Fig. 5A). Apparently, this is not enough to sufficiently suppress replicase, and upon further evolution C1732U is selected. This mutation, which we call T_2, stabilizes the terminator and thus contributes to replicase down regulation. In Fig. 5A we show the resulting revertant, revBVT$_2$.

Another way to improve on revB is A1744C (called here M_3) (Fig. 5, center). Although this change seemingly destabilizes MJ further, a closer look shows that it promotes an alternative pairing (Fig. 5C; revBM$_3$). We suppose that the advantage of this new folding is that the 4-nt gap has now seemingly disappeared and there is again the 1-nt distance between the operator and MJ. Further evolution of revBM$_3$ shows the well-known mutation V (U1817C), indicating that A1744C (M$_3$) was not enough to silence replicase to the appropriate level.

RevBM$_3$ can also occur in combination with mutation M_2 (C1743U) (Fig. 5, center). Again, this change may seem to destabilize MJ further, but in fact it leads to a reshuffling in the MJ pairing such that the 4-nt gap is seemingly reduced to 3 nt, and the distance between operator and MJ again back to wt (revBM$_2$M$_3$; Fig. 5D).

A fifth way to improve on revB is the change A1731C (mutation T_1), which we have also seen in one of the revC descendants (Fig. 5, center). At first glance the mutation seems to further destabilize the terminator hairpin. However, the data can be understood by recognizing that there are two possible versions of the RNA folding in mutant X (Fig. 1C). In structure II the MJ interaction is intact, with 7 base pairs, and in structure I the 4-nt deletion is divided over MJ and hairpin T, both lacking 2 base pairs. Base pair stacking energies predict structure I to be slightly more stable and therefore mostly being formed, in line with the observation that the 4-nt deletion in mutant X causes a big rise in uncoupled replicase synthesis (Fig. 4). The only way to force formation of structure II is to make G1716 and G1717 pair with something else rather than with C1734 and C1735. This is exactly what mutation A1731C will do because it favors an

alternative folding for hairpin T (Fig. 5B). Now, the MJ structure looks like the wt except that C1714 does not have a pairing partner. To verify that the A1731C suppressor mutation is in fact selected because it restores control, we compared the leaky replicase translation in revB with that in revBT$_1$. Indeed, Fig. 4 shows that leaky replicase synthesis in revBT$_1$ is reduced five times compared to that in revB (227 versus 1070 units) and is back to the wt level or nearly so.

Competing Optimized Versions of RevB and RevC

Table 1 shows that revB quickly outgrows revC. The question that remains is whether revC, optimized in the absence of revB, can compete with revB and its improved progeny. It is conceivable that a revertant, once having taken a wrong path, will never be able to improve itself to the level of other revertants that have initially taken a better road.

We thus competed improved versions of revB, i.e., revBV and revBT$_1$, against the improved revC versions revUV, revCCU, and revACU (Figs. 3, 5, and 9). RevUV looses from revBV and revBT$_1$. This is concluded from the observation that in the sequence gel on total RNA, that of revUV was no longer seen. RevCCU and revACU, on the other hand, turn out to be about equally as good as either revBV or revBT$_1$. It is clear, then, that even if initially evolution follows an inferior path, later restorative mutations can make up for the arrears. Of course, this may not be a general rule. Our results simply show that it is possible. The chance nature of the mutational process is probably also responsible for our finding that revB does not always drive out revC at the same pace. Sometimes, revB and revC progeny coexist for some time. We suppose that this can only happen if revC has improved itself before being outnumbered by revB and therefore survives longer than on average. The point is illustrated by analysis 10.8 (Table 1), where improved versions of revC were indeed found in the virus mixture. Normally, revC constitutes the majority of the population at cycle 2 (as revealed by lysate sequencing) but is completely outgrown by unimproved revB at cycle 5 (analysis 1.5; Table 1).

RevL, RevS, and RevD

RevL was found as one of the 40 sequenced plaques. It has an extra A in the loop of the operator, which is predicted to destroy its coat-protein binding capacity (Fig. 6B, left). It seemed interesting to study its evolution because unlike revA and revC, revL may have a poor mutational neighborhood (Burch and Chao 2000); that is, there is no plausible pathway leading quickly to either the F6 aptamer or to a tetraloop with A residues in the first and fourth position (see Dis-
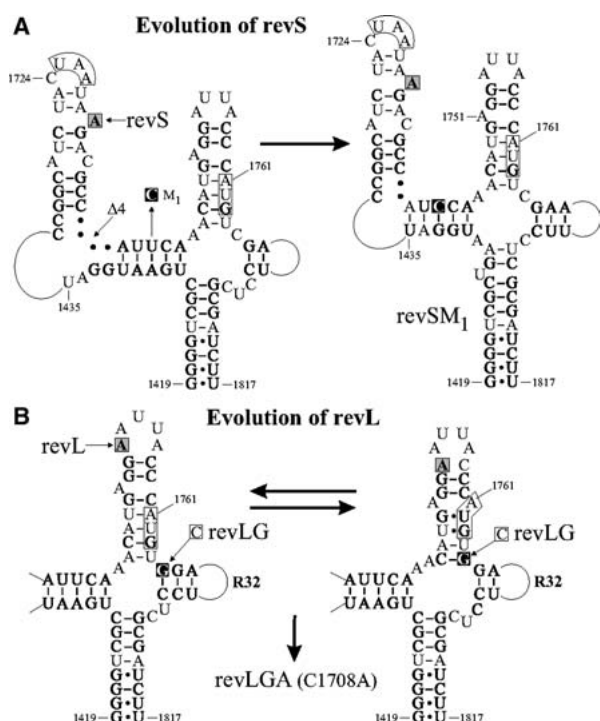
Fig. 6. A Evolution of revS. The left panel shows suppressor mutation U1742C (M_1) in the original RNA structure; the right panel, in the context of the induced structure. Note the similarity in MJ pairing to revBM$_3$V, despite the presence of the extra A in revBM$_3$V. The two dots in the right panel are there to illustrate that only 2 nt are lacking to form a structure that looks like the wild type. **B** Evolution of revL. Both structures shown are predicted to diminish translation of the replicase gene, both coupled and uncoupled. The coupling itself is predicted not to be affected. In later stages C1708A is selected but we cannot interpret this substitution.

cussion). Indeed, evolution of revL does not produce a recognizable operator hairpin. Instead, after eight passages we obtained revLG with the A1765G change. Two interpretations are possible based on two different foldings. In one (Fig. 6B, left) hairpin R32 is stabilized, which is expected to decrease leakage of replicase production. Nothing is done, however, to repair the operator. In the second interpretation the A1765G change is thought to stabilize the alternative structure shown in Figure 6B (right). Possibly, this stem-loop may serve as a surrogate operator. The puzzling aspect here is that one would expect to find A1757U or A1753U to create a triloop with the bulged A in the correct position. One still would need then U1756A to arrive at the F6 aptamer. (A third explanation could be that the codon change to tryptophan somehow contributes to replicase repression.)

Upon further passaging the mutation C1708A developed (revLGA), but we cannot explain its selection. In our structure model C1708 is an unpaired nt (Fig. 7B). RevL was evolved again in the expectation to obtain other solutions, yet the same revertant, revLG, was obtained. This demonstrates that this mutant is unable to produce further adaptive mutations and is therefor quickly lost during propagation in a large population size. It is not clear why revL does not select mutations like V, T_1 or T_2 that stabilize the MJ pairing.

RevS surfaced one time after passaging an amount of lysate corresponding to about 50 pfu (Table 1, analysis 2.4). It has a one nt insertion in the stem of the terminator hairpin T (Fig. 6A). The operator is intact but as with all others the MJ interaction is defect. RevS was evolved from an isolated plaque. It selected the U1742C change (mutation M_1), which seemingly further destabilizes MJ. However, this revertant, revSM$_1$, resembles revBM$_3$, as it can form a similar MJ folding (Fig. 6A, right). We notice a preference to involve G1432 and G1433 in the MJ pairing, possibly because of the great stability these neighboring G-C pairs provide.

To find other solutions revS was evolved a second time but the same revertant was found. RevS appeared only once. Its fitness or its probability of formation must therefore be relatively low, and its one-time appearance in bulk evolution a chance event. As can be seen from analysis 2.4 (Table 1), revS is ultimately replaced by a revB descendent.

RevD was found as one of the plaques. Its evolution was not studied.

RevI4

RevI4 has a CAAA duplication between MJ structure and operator (Figs. 2B and 7A). The insertion does not restore the MJ interaction and further evolution of an I4 plaque showed the well-known A1731C suppressor mutation which rearranges the T hairpin and stabilizes the MJ interaction (revI4T$_1$). In fact, we expected to find as revertant the "belied prediction" shown in Fig. 7C, because here the 4-nt insertion is used to return to the wt RNA structure. In trying to get this result revI4 was evolved a second time using larger amounts of phages for transfer. Applying this protocol, we obtained revI4U with the C1439U change, which remains unexplained (Fig. 7B). Further passaging produced revI4U-I3, which has a 3-nt GCC duplication that almost filled the gap present in mutX. The insert restores the MJ interaction. Nevertheless, our prediction for the evolution of I4 (Fig. 7C) did not come true. This illustrates how the mutational bias toward frequent but inferior solutions can frustrate the potential to reach a superior solution on an alternative adaptive peak (our belied prediction).

RevI4 is not a particularly good revertant. It showed up only once and was, after 10 cycles, displaced by a revB derivative (Table 1).

RevIN1 and RevIN10

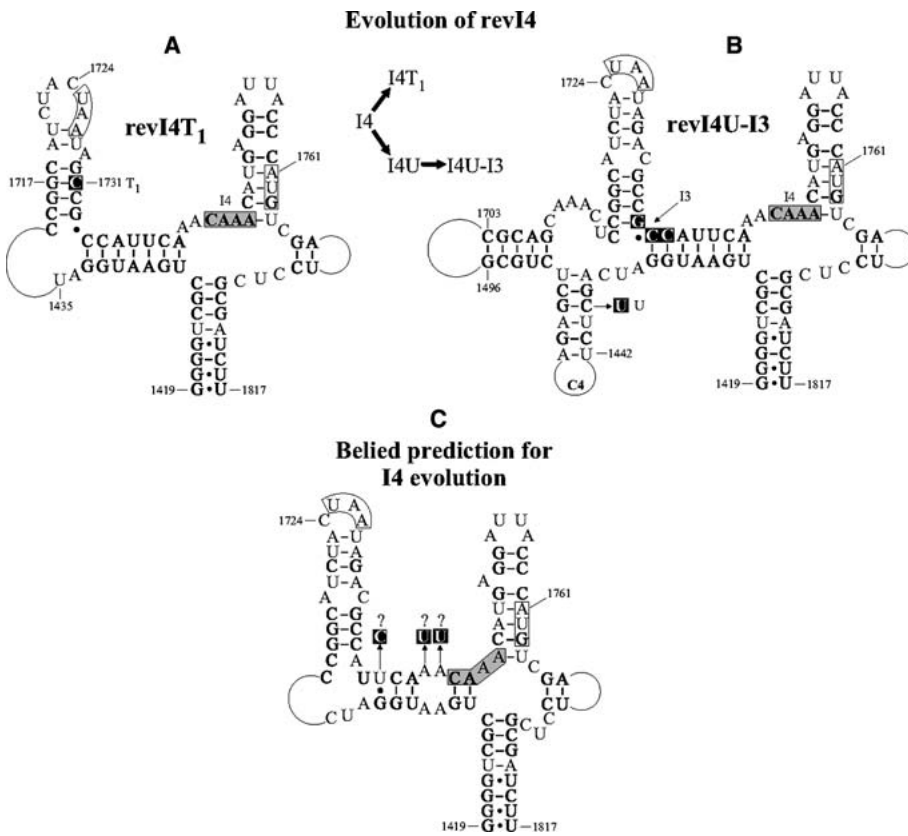Further increments in the sample size for evolution cycle 2, to 400 pfu, revealed two new revertants,

Evolution of revI4

Fig. 7. Evolution of revI4. **A** RevI4T$_1$ carries suppressor mutation T$_1$, which we have seen and discussed before. **B** In the other pathway, revI4U, the selection of C1439U is uninterpretable but the subsequent GCC duplication revI4U-I3 is self-explanatory. **C** The "educated guess" of the authors about the path evolution of revI4 was expected to take. We ran evolution of revI4 two times but our prediction was belied.

revIN1 and revIN10. RevIN1 contains a single C insertion at the site of the original GGCC deletion, where it adds again one base pair to the weakened MJ interaction (Fig. 2A). We have not found improved versions of revIN1, not even after 20 cycles. RevIN10 showed a 10-nt duplication covering the site of the deletion (Fig. 2B). Its MJ structure can be fully restored, and compared to wt it has six redundant nt between the terminator and the MJ pairing (not shown). These extra bases seem not to be a big burden, as continued evolution did not show any deletions or other changes. RevIN1 and revIN10 turn up frequently as winners (Table 1).

RevIN4 and RevIN7

The last two new revertants became visible only after the size of the lysate used for the second evolution cycle was increased to about 10^5 pfu. RevIN7 has a 7-nt duplication, which can completely restore MJ (Fig. 2B). Its difference from wt is a 3-nt insertion between terminator and MJ. Upon further passaging it selects an unexplained G → A substitution in the duplication (GCCAU to ACCAU, indicated as IN7A in Fig. 9).

RevIN4 has a 4-nt AUUC duplication precisely at the site of the initial deletion (Figs. 2B and 8). The duplication almost fully restores MJ, except that it has a U*G pair where the wt has C-G. Furthermore, the two lower base pairs of the terminator helix
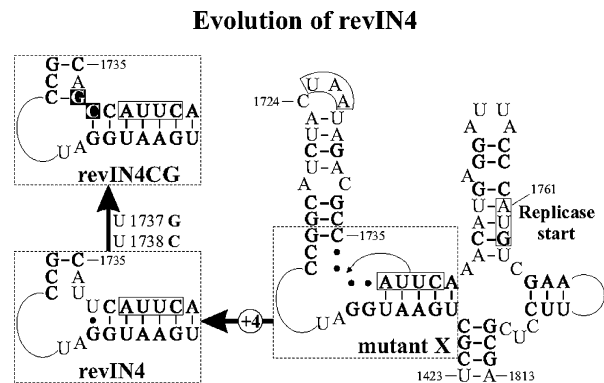
Evolution of revIN4

Fig. 8. Evolution of revIN4. This is the only return to wild type that we encountered in the whole study. RevIN4CG descended from revIN4C, not in the bulk evolutions presented in Table 1 but in another passaging.

T have become mismatches. Upon further evolution of revIN4, first there is the U1738C change that brings the MJ pairing back to wt (revIN4C). Then we find U1737G, a change that restores the lower base pair in the terminator hairpin. The surprising fact now is that this revertant is only one nt away from the wt sequence (we have not allowed this last step to occur because we cannot verify, when finding the wt, that it is not a contamination).

Not surprisingly, revIN4 (and progeny) is the best revertant found in this study. It has recovered the four lost nt in the right place. From there on, it is a matter of two transitions and one transversion to
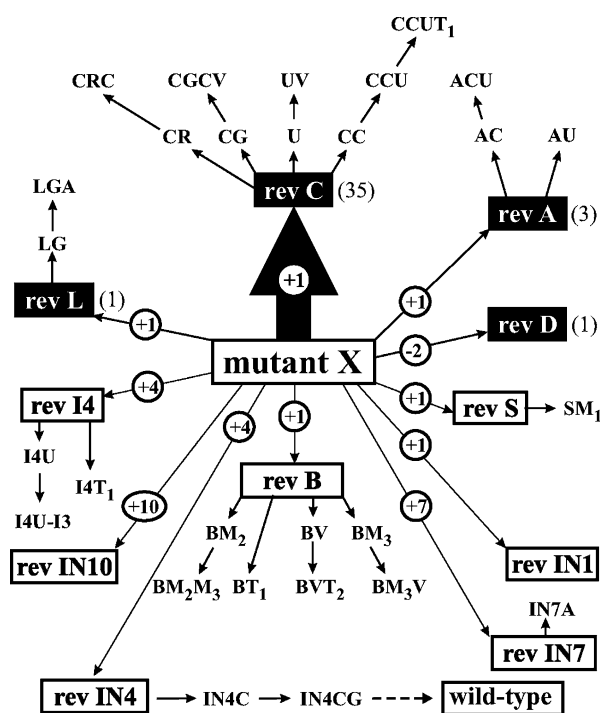
Fig. 9. Overview of the revertants obtained in this study. Black boxes show revertants identified in plaques. In parentheses, their number, found after analysis of 40 plaques, is shown. White boxes show primary revertants (one mutation) obtained by bulk evolution. Arrows leading away from the boxes show the further progeny. The number of deleted or inserted nucleotides is encircled.

arrive at the wt sequence. RevI4 also recovered four nt but these were in the wrong place, and as we have seen the phage was unable to select the substitutions (Figs. 7A and B) that would have led back to wt (Fig. 7C). Instead, the RNA got entangled in alternative structures like revI4T$_1$ and revI4U-I3 (Figs. 7A and B).

It is good to realize that none of the revertants made here and in our previous studies can stand up to the wt. This is not necessarily due to any obvious defect in a structure element. Rather, slight differences in hairpin stability can already result in severe losses in fitness (Olsthoorn et al. 1994). RevIN4C, for example, differs from wt only in positions 1736 and 1737, resulting in a weaker terminator stability, while it has two mismatches at the bottom of this hairpin. They were mixed at a 1:1 ratio and grown. After three cycles the ratio was 10:1, after four cycles 20:1, in favor of wt.

Discussion

Relative Fitness of the 11 Revertants Isolated

Here we uncover the multiple ways in which the ssRNA phage MS2 repairs a 4-nt deletion that disrupts both the reading frame of the lysis gene and the control of replicase translation. Two complementary

methods were used to obtain a large number of different revertants at a reasonable effort. High-frequency solutions were uncovered by plating the lysate of an infected bacterial culture. In 40 sequenced plaques we found four different solutions, 35 plaques being identical (Fig. 9). Less frequent but higher-fitness solutions could be obtained by passaging a lysate sample a few times to enrich for the low-abundance but high-fitness revertants. By taking larger and larger lysate samples for the first enrichment step, one will select for gradually higher and higher fitness. This approach, which is limited by the volume of lysate one can process, yielded seven new solutions, bringing the total amount of analyzed revertants to 11. It follows from the method of selection that the solutions obtained by plating are least fit. Solutions worse than those obtained by plating are bound to exist but cannot be selected for, and they can only be found by analyzing large amounts of plaques obtained from the first lysate and comparing them with revC.

Among the plaques, revA and revL are probably the least fit revertants since, in contrast to revC and revD, we do not see a trace of them after five cycles (first several rows in Table 1). However, we cannot exclude that the dominant presence of revC is a result of its overrepresentation in the starting sample. At the upper end of the scale we find the fittest revertant revIN4, emerging at the largest evolved lysate volume and evolving to wt.

Order of Repair of the Damage

The initial deletion of four nt causes two defects: frame shift and loss of replicase control. This study shows that the more serious ailment, loss of the lysis protein, is repaired first. Evidently, this apparent order results from the fact that those revertants that have repaired the worst defect outgrow those that have fixed a small defect. Therefore, the order in which we see repair of the various handicaps is also the order in which these handicaps contribute to fitness.

In revC and revA the frame shift is repaired by the insertion of an extra nt into the operator hairpin, causing its inactivation. This is an unfortunate coincidence for the phage and an example of how a restorative mutation can be selected for even if it has a strong pleiotropic effect on other functions. Now, there are again two defects, operator and replicase control. Upon evolution of revC and revA we see that in the majority of cases (revCR being the exception), rebuilding of an operator has priority. This is accomplished as described under Results. Interestingly, the phage was able to construct a new type of operator that we do not find in nature but that was nearly identical to one obtained previously via an in

vitro selection procedure (Convery et al. 1998). The next and final step concerns the repair of replicase control.

Various Ways to Restore Replicase Control

This control is the result of five structures that cooperate in keeping the replicase start site blocked in the absence of coat-gene translation. These structures are terminator hairpin T, the MJ pairing, the operator, hairpin R32, and structure VD. Structures T and MJ are strongly weakened by the deletion. There are many different ways to restore this control. As the five structures work together, a defect in one, MJ, may be compensated by extra stability in another. One example is mutation V (U1817C) stabilizing the VD pairing. Another is mutation T_2 (C1732U) stabilizing the terminator. Then there is mutation R (A1768G) probably stabilizing hairpin R32.

A further possibility is to reinforce MJ directly by selecting base substitutions that allow an alternative and presumably stronger pairing. The examples here are mutations M_2 and M_3 (in revB) and mutation M_1 in revS. A final solution for the MJ defect is mutation T_1 (A1731C) causing a rearrangement in the terminator stem, leading to a wt MJ pairing.

Comparing the Fitness of RevC, RevB, and RevIN1

From their appearance and disappearance outlined in Table 1, we can infer that revIN1 is better than revB, which in turn is better than revC. For these three revertants it is easy to understand the ranking. RevC restores the frame shift at the expense of the operator and thus exemplifies the principle of two steps forward, one step back (Otto 2003). Sticking to the same metaphor, we could say that revB does two steps forward and none back. It only repairs the frame shift. RevIN1, on the other hand, does three steps forward. Not only does it fix the frame shift, but also it aids the MJ structure with one extra base pair. The resulting structure is apparently so satisfactory that further suppressor mutations, such as, e.g., C1732U or U1817C, were not found. The remaining 3-nt gap in revIN1 could also have been filled by a 3-nt duplication as seen in I4U-I3. As this possibility is not employed, we must assume that the phenotype of revIN1 is already so close to wt that further improvements are contributing too little to outgrow their parent in the time span of the experiment. Alternatively, and more likely, the 3-nt duplication has too low a probability to take place.

RevIN7 and RevIN10

These revertants insert 7 and 10 nt precisely at the site of the deletion and the original MJ pairing is recovered. In this sense these revertants are one step ahead of revIN1 since they do not suffer from the small defect in MJ pairing. RevIN7 and revIN10 only have to delete three and six nt, respectively, to be one nt away from wt. However, over the duration of the experiment it did not happen.

RevIN4 and RevI4

RevIN4 is the superior revertant, as it replaces the 4 missing nt with new ones and it does so precisely at the site of the deletion. It differs in only 3 nt from the wt sequence (Fig. 2). Still, if the 3 mutant nt would have caused an alternative local folding, the subsequent restorative mutations would have been selected for the extent to which they would have suppressed replicase synthesis by stabilizing the alternative structure, and return to wt would have been cut off. However, in the present case revIN4 seems to adopt a wt structure and can therefore evolve to wt (in fact only the two lower base pairs of the terminator are missing).

RevI4 also has a 4-nt insertion, but at a different site. Here, it has apparently been impossible to return to the wt structure. This is illustrated in Fig. 7B, where we draw the pairing scheme that needs to be adopted to return to the wt. One would need at least the two transversions A1741U and A1742U to form a stable MJ pairing. The same or maybe even a better MJ pairing can be obtained by the single A1731C mutation and we can thus rationalize why we obtain this revertant rather than any derivative of the belied prediction shown in Fig. 7C. It is clear, then, that by the patchwork solution of A1731C the opportunity to exploit the inserted CAAA sequence to return to the wt structure has vanished.

RevS and RevL

RevS and revL are two more examples that probably any single-letter insertion will do to obtain a viable revertant. In revS this results in a bulge in the upper part of the terminator helix. This structure is somewhat reminiscent of revBT₁. To reinforce the MJ structure the mutation U1742C is selected, leading to an alternative MJ pairing and an apparent gap of two rather than four nt (revSM₁; Fig. 6). One might have expected the additional mutation U1817C, but maybe as it is the structure is strong enough.

RevL has an extra A in the loop of the operator and resembles revC and revA in that all three have an inactive operator with a loop of five nt. But whereas revC and revA can easily escape to a loop of either three or four nt with an A in the 3′ loop position, the possibilities for revL are severely limited. It can escape to a loop with four nt by the A1753G change

but this is a mutation in the Shine Dalgarno sequence of the replicase gene. Such changes have profound effects on translation efficiencies. RevL could also develop a loop of three by an A-to-U transversion but the third position of the loop would be a U rather than the A, which is important for interaction with the coat protein. In other words, revL has a poor evolutionary neighborhood (Burch and Chao 2000). The revertant selects C1765G, a choice which we have tentatively tried to explain in Fig. 6. Particularly puzzling is that this revertant does not seem to compensate for the feeble MJ structure. It could have done so in the various ways we have discussed above.

Sequence of Events in Adaptation

We have discussed above that the most serious defect seems to be repaired first. Most likely, this is the result of phenotypic selection where those genomes that benefit most from their adaptive mutation outgrow those that benefit less.

Another question is whether the first adaptive mutation sets the stage for the ones to come or whether all restorative mutations are independent of each other and can appear in random order. The first scenario is more plausible. In a general sense one must consider that each restorative mutation is a response to the prevailing situation (the phenotype). Each subsequent mutation must take into account what the new reality is that is created by the previous mutation. In this study it means many times either that the first adaptive mutation begins to optimize an alternative structure and that this path must now be followed or that the first adaptive mutation solves one problem but creates a new one that must be fixed by the mutations to follow. For example, revC and revA restore lysis at the expense of the operator and the next mutation must deal with that fact. As it happens, this can be done in several ways (Fig. 5) but the fact remains that mutation 2 is a response to the situation brought about by mutation 1. Another example is revCG, where the operator is saved by the creation of a G-C pair, which now leads to the formation of a triloop and the F6 aptamer (Fig. 3B). However, this G-C pair makes the operator too stable to permit replicase translation, and indeed the next mutation destabilizes the stem by dissolving the lower base pair (Fig. 3B; revCGC). This has been observed in earlier studies of the adaptation of RNA phage in which the thermodynamic stability of an important helix was compromised. Upon evolution the same stability was developed, albeit with the use of different base pairs. During the process there was many times an overshoot in the effort to reach the wt stability, which was then corrected by the subsequent mutation (Olsthoorn et al. 1994).

Another illustration of the way structural constraints determine the order and choice of suppressor mutations is provided by the restoration of replicase control. Broadly speaking there are three ways to achieve this: first, mutations in the 3′-sequence of MJ that enable a stronger but alternative pairing mode, e.g., mutations M_1, M_2, and M_3; second, mutations that sacrifice the terminator hairpin to restore MJ (mutation T_1); and third, mutations that do not induce alternative foldings but stabilize surrounding helices (mutations T_2 and V). Here, we learn that mutations can exclude others from being selected. For example, mutation T_1 rebuilds MJ in a specific way and thereby excludes the selection of M_1, M_2, or M_3, and vice versa. Similarly, mutation T_2 stabilizes the terminator and will be incompatible with mutation T_1, and vice versa. On the other hand, mutation V stabilizes the VD pairing somewhat. The VD structure itself is always present in our experiments and the U1817C mutation does not interfere with any other structure and can always be adopted to fine-tune replicase expression.

Origin of Replication Errors

Nucleotide substitutions and, in particular, transitions are the most frequent errors made by viral RNA replicases ($\sim 10^{-4}$/nt). Insertions and deletions are much less likely, and in this category slipping forward or backward over a row of identical bases is the dominant mistake. Comparing the titers of mutant X and the wt, we estimate the probability of inserting the extra nt in revC to be 10^{-6} per replication. For revB, which occurs roughly 100 times less frequently than revC, this probability should be about 10^{-8}. We suppose that revC occurs more frequently than revB because the newly synthesized strand, even when slipped, sticks better with C-G than with A-U pairs. Duplication of what seem to be random sequences (revIN4, revIN7, and revIN10) is even less likely than slippage over a row of identical nucleotides. From their frequency of occurrence (first seen in samples of 10^5 pfu), we estimate a probability of 10^{-10} per replication for revIN7 and revIN4 to be formed. If the probability of repairing a defect times the number of phages present in the passage volume is <1, the phage can be considered dead.

It is interesting that three revertants (revA, revL, and revS) have acquired an extra A that is not, as in revB, provoked by a row of A's (Fig. 2A). We can think of two explanations. First, it is known that RNA replicases like Qβ replicase will add an A to the chain when idling at the end of the template. If the enzyme copies a broken minus strand, it may thus add an A to a broken plus strand. Subsequent RNA recombination (White and Morris 1995) may then

introduce the extra A in the complete genome. The other explanation is the addition of an A by the host enzyme poly(A) polymerase. Such additions have been shown to occur (van Meerten et al. 1999). Here the sequence of events is cleavage of MS2 RNA by a host endonuclease, followed by polyadenylation of the newly created 3′ end by poly(A) polymerase. Thereafter the exonuclease RNaseII and/or polynucleotide phosphorylase will degrade the RNA starting at the poly(A) tail. However, such MS2 RNA fragments can still take part in RNA recombination. As a result, the A (or A's) added by a host enzyme can become part of the phage genome (van Meerten et al. 1999).

Evolutionary Opportunities

The evolutionary perspectives for the various primary revertants are quite different. In particular, RevL and revI4 seem to have a poor mutational neighborhood, and as a result, they fail to develop progeny that stands up in the bulk evolution experiments. On the other hand, there is revB with at least four different pathways to promising descendants (Fig. 9). As shown in Table 1, revB descendants keep dominating the progeny as long as the samples are small enough to exclude the appearance of super solutions like revIN10, revIN7, and revIN4. (Interestingly, IN10 and IN7 show only slow or no further adaptation. Perhaps, their high fitness makes it difficult for further suppressor mutations to manifest themselves in the short time scale of our experiments.) RevC also seems to have a rich evolutionary neighborhood. Figure 9 shows that there are at least four different possibilities for further development.

References

Berkhout B (1985) Translational control mechanisms in RNA bacteriophage MS2. PhD thesis, Leiden University

Berkhout B, de Smit MH, Spanjaard T, Blom T, van Duin J (1985) The amino terminal half of the MS2-coded lysis protein is dispensable for function: implications for our understanding of coding region overlaps. EMBO J 4:3315–3320

Berkhout B, van Duin J (1985) Mechanism of translational coupling between coat protein and replicase genes of RNA bacteriophage MS2. Nucleic Acids Res 13:6955–6967

Burch CL, Chao L (2000) Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature 406:625–628

Convery MA, Rowsell S, Stonehouse NJ, Ellington AD, Hirao I, Murray JB, Peabody DS, Phillips SEV, Stockley PG (1998) Crystal structure of an RNA aptamer-protein complex at 2.8 A resolution. Nat Struct Biol 5:133–139

de Smit MH (1998) Translational control by mRNA structure in eubacteria: molecular biology and physical chemistry. In: Simons RW, Grunberg Manago M (eds) RNA structure and function. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 495–540

de Smit MH, van Duin J (1990a) Control of prokaryotic translational initiation by mRNA secondary structure. Prog Nucleic Acid Res Mol Biol 38:1–35

de Smit MH, van Duin J (1990b) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci USA 87:7668–7672

Drake JW (1993) Rates of spontaneous mutation among RNA viruses. Proc Natl Acad Sci USA 90:4171–4175

Gonzalez de Valdivia EI, Isaksson LA (2004) A codon window in mRNA downstream of the initiation codon where NGG codons give strongly reduced gene expression in Escherichia coli. Nucleic Acids Res 32:5198–5205

Klovins J, van Duin J (1999) A long-range pseudoknot in Qβ RNA is essential for replication. J Mol Biol 294:875–884

Klovins J, Tsareva NA, de Smit MH, Berzins V, van Duin J (1997a) Rapid evolution of translational control mechanisms in RNA genomes. J Mol Biol 265:372–384

Klovins J, van Duin J, Olsthoorn RCL (1997b) Rescue of the RNA phage genome from RNase III cleavage. Nucleic Acids Res 25:4201–4208

Klovins J, Berzins V, van Duin J (1998) A long-range interaction in Qβ RNA that bridges the thousand nucleotides between the M-site and the 3′ end is required for replication. RNA 4:948–957

Licis N, van Duin J, Balklava Z, Berzins V (1998) Long-range translational coupling in single-stranded RNA bacteriophages: an evolutionary analysis. Nucleic Acids Res 26:3242–3246

Licis N, Balklava Z, van Duin J (2000) Forced retroevolution of an RNA bacteriophage. Virology 271:298–306

Lim F, Downey TP, Peabody DS (2001) Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7. J Biol Chem 276:22507–22513

Lowary PT, Uhlenbeck OC (1987) An RNA mutation that increases the affinity of an RNA-protein interaction. Nucleic Acids Res 15:10483–10493

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

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

Olsthoorn RCL, van Duin J (1996a) Random removal of inserts from an RNA genome: selection against single-stranded RNA. J Virol 70:729–736

Olsthoorn RCL, van Duin J (1996b) Evolutionary reconstruction of a hairpin deleted from the genome of an RNA virus. Proc Natl Acad Sci USA 93:12256–12261

Otto SP (2003) Two steps forward, one step back: the pleiotropic effects of favored alleles. Proc R Soc Lond B 271:705–714

Remaut E, Stanssens P, Fiers W (1981) Plasmid vectors for high-efficiency expression controlled by the P_L promoter of coliphage lambda. Gene 15:81–93

Rosenberg AH, Goldman E, Dunn JJ, Studier FW, Zubay G (1993) Effects of consecutive AGG codons on translation in Escherichia coli, demonstrated with a versatile codon test system. J Bacteriol 175:716–722

Taniguchi T, Palmieri M, Weissmann C (1978) Qβ DNA-containing hybrid plasmids giving rise to Qβ phage formation in the bacterial host. Nature 274:223–228

Tars K, Fridborg K, Bundule M, Liljas L (2000) The three-dimensional structure of bacteriophage PP7 from Pseudomonas aeruginosa at 3.7-A resolution. Virology 272:331–337

van den Worm SH (2004) RNA-protein interactions in the ssRNA phages. PhD thesis, Leiden University

van Duin J (1999) Single-stranded RNA phages (Leviviridae). In: Webster RG, Granoff A (eds) Encyclopedia of virology. Academic Press, London, pp 1663–1668

van Himbergen J, van Geffen B, van Duin J (1993) Translational control by a long range RNA-RNA interaction; a basepair substitution analysis. Nucleic Acids Res 21:1713–1717

van Meerten D,Zelwer M, Regnier P, van Duin J (1999) *In vivo* oligo(A) insertions in phage MS2: role of *Escherichia coli* poly(A) polymerase. Nucleic Acids Res 27:3891–3898

van Meerten D, Girard G, van Duin J (2001) Translational control by delayed RNA folding: identification of the kinetic trap. RNA 7:483–494

White KA, Morris TJ (1995) RNA determinants of junction site selection in RNA virus recombinants and defective interfering RNAs. RNA 1:1029–1040

Witherell GW, Gott JM, Uhlenbeck OC (1991) Specific interaction between RNA phage coat proteins and RNA. Prog Nucleic Acid Res Mol Biol 40:185–220