# Leeway and constraints in the forced evolution of a regulatory RNA helix

Olsthoorn, R.R.C.L.; Licis, N.; Duin, J. van

The EMBO Journal vol.13 no.11 pp.2660–2668, 1994

# Leeway and constraints in the forced evolution of a regulatory RNA helix

**R.C.L.Olsthoorn, N.Licis[1] and J.van Duin[2]**

Department of Biochemistry, Leiden University, Gorlaeus Laboratories, PO Box 9502, 2300 RA Leiden, The Netherlands
[1]Present address: Institute of Molecular Biology, Latvian Academy of Sciences, Krustpils 53, Riga 226065, Latvia
[2]Corresponding author

Communicated by P.van de Putte

**The start of the coat protein gene of RNA phage MS2 adopts a well-defined hairpin structure of 12 bp (including one mismatch) in which the start codon occupies the loop position. An earlier expression study using partial MS2 cDNA clones had indicated that the stability of this hairpin is important for gene expression. For every −1.4 kcal/mol increase in stability a 10-fold reduction in coat protein was obtained. Destabilizations beyond the wild-type value did not affect expression. These results suggested that the hairpin was tuned in the sense that it has the highest stability still compatible with maximal ribosome loading. Employing an infectious MS2 cDNA clone, we have now tested the prediction that the ΔG° of the coat protein initiator helix is set at a precise value. We have introduced stabilizing and destabilizing mutations into this hairpin in the intact phage and monitored their evolution to viable species. By compensatory mutations, both types of mutants quickly revert along various pathways to wild-type stability, but not to wild-type sequence. As a rule the second-site mutations do not change the encoded amino acids or the Shine–Dalgarno sequence. The return of too strong hairpins to wild-type stability can be understood from the need to produce adequate supplies of coat protein. The return of unstable hairpins to wild-type stability is not self-evident and is presently not understood. The revertants provide an evolutionary landscape of slightly suboptimal phages, that were stable at least for the duration of the experiment (~20 infection cycles). The results also show that translation-initiation frequency does not appear to depend on the primary sequence (except for the start codon and the Shine–Dalgarno region) but solely on the ΔG° of the structure that encompasses the ribosomal binding site.**

*Key words:* molecular evolution/RNA phage/RNA structure/translation initiation

## Introduction

The RNA of the single-stranded RNA phage MS2 is ~3500 nt long and contains four genes. Their products contribute to phage maturation, encapsidation, lysis of the host and RNA replication, respectively (Figure 1a). The coat protein, which is made in large amounts, has a dual role. It encapsidates the RNA, but it also acts as a translational repressor by binding to a hairpin that contains the beginning of the replicase gene (Witherell *et al.*, 1991). This last property limits replicase synthesis to a small burst early in the infection cycle.

MS2 RNA, like the other phage RNAs, adopts a unique secondary structure that is assumed to be important in many aspects of the life cycle (Skripkin *et al.*, 1990; Skripkin and Jacobson, 1993). It condenses the RNA, allowing it to be packaged in the virion. At the same time the formation of stem–loop structures is thought to be necessary to separate mother and daughter strands during replication (Mills *et al.*, 1978; Priano *et al.*, 1987) and to protect against cellular RNases. In addition, the presence of base-paired structures in coding regions may mediate ribosome pausing, which could help in the correct folding of the nascent protein (Varenne *et al.*, 1984; Guisez *et al.*, 1993). Such assumptions are quite reasonable, but not easy to test. In fact, only for a very small number of structures has a function been identified by measuring gene expression in response to changes in RNA structure. Using this approach stem structures at the beginning of the maturation, lysis and replicase genes were shown to exert negative control over ribosome access to these start sites (van Duin, 1988; H.Groeneveld and J.van Duin, in preparation). In addition, a long range RNA–RNA interaction was shown to be responsible for translational coupling between the coat and

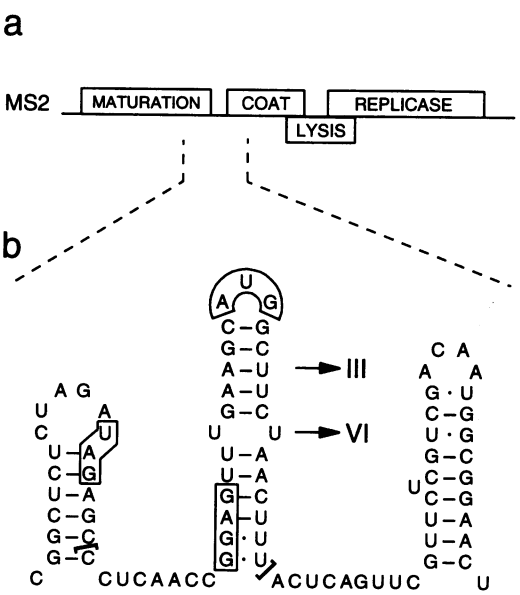


**Fig. 1.** (**a**) Map of the MS2 RNA genome. (**b**) Secondary structure at the start of the coat protein gene. Its initiation codon, the Shine–Dalgarno sequence and the UAG stop codon of the poorly translated maturation protein gene are boxed. Roman numerals are used to number the base-pairs in the stem. Brackets indicate the RNA region protected against pancreatic RNase in a translational initiation complex (Steitz, 1969).

replicase genes (van Himbergen *et al.*, 1993). In $Q_\beta$ RNA the binding sites for $Q_\beta$ replicase in plus and minus strands have been identified, but no experiments have been done to relate this binding to the secondary structure of the RNA region involved (Meyer *et al.*, 1981; Barrera *et al.*, 1993).

In this paper we analyze the function of the structure present at the start of the coat gene. This region is shown in Figure 1b. It adopts a simple hairpin structure with the start codon in the loop and the Shine–Dalgarno (SD) region at the base of the stem. Its calculated stability at 37°C is −7.0 kcal/mol. In a previous study using partial MS2 cDNA clones, we constructed some 20 different helix mutants in which the stability was varied from −3.0 to −11.0 kcal/mol, without changing the SD complementarity or the encoded amino acids. Measuring coat protein production by these mutants revealed that base changes that strengthen this hairpin have a drastic negative effect on translation. Each increase in stability of −1.4 kcal/mol causes translation to fall by a factor of 10. For instance, turning the U·U mismatch in the middle of the stem into either a U-A or an A-U match increases the stability by 3.5 kcal/mol and consequently the yield of coat protein drops to 0.2% of the wild-type value in either mutant (de Smit and van Duin, 1990).

Surprisingly, further destabilization of the stem beyond the wild-type value did not lead to increased coat protein yields. Instead, the amounts stayed at the wild-type level. Mathematical analysis of the data revealed that ribosomes only bind the unfolded state of the helix. Furthermore, the wild-type stem appeared just weak enough to allow unabated translation and the data therefore suggested that the hairpin has evolved to have the highest stability still compatible with maximal ribosome loading.

To examine the presumed tuning of this helix in real life, we introduced versions of this hairpin with differing stability into an infectious clone of full-length MS2 cDNA. The resulting mutant phage lysates were then used to infect fresh $F^+$ host cells. The evolving sequences of the descendants were monitored after various (up to 25) rounds of infection. Stabilized as well as destabilized hairpin variants reverted to a pseudo wild-type after a few cycles, i.e. the hairpin stability in all revertants returned stepwise to that of the wild-type, without adopting the wild-type sequence. This is achieved by second-site mutations that compensate for the inflicted loss or gain in helix stability. Once wild-type stability is reached the sequence becomes stable, at least for the duration of the experiment. The results show that this MS2 helix is strongly conserved and it must therefore have an important function. In addition, for a virus to be viable, the stability needs to be maintained within narrow limits. Selection against high stabilities is obviously related to translational efficiency, but the disadvantage of low stability remains unknown.

## Results

### *The system*

The complete MS2 cDNA was cloned in a plasmid under control of the $p_L$ promoter of phage λ in strain M5219 as described in Materials and methods. Such plasmids are potentially lethal to the cell since they produce phage spontaneously, even when the $p_L$ promoter is repressed. Furthermore, mutants that produce very little coat protein have two additional problems in producing viable phage. First, they cannot encapsidate their RNA, a condition that endangered our planned experiments to produce mutant phage. Secondly, they cannot adequately repress replicase synthesis. We and others have found previously that the replicase protein is lethal to the cell (Remaut *et al.*, 1982; our unpublished observation). To stabilize the infectious clone and the derived mutants, a second plasmid was introduced carrying the coat protein sequence inserted in the *tet* gene of pACYC184 and which we call pCOAT184. This plasmid constitutively synthesizes coat protein at a level sufficient to repress replicase translation by ~90% (Berkhout, 1986). Indeed, in the absence of pCOAT184, many plasmids, in particular those predicted to produce low amounts of coat protein, frequently suffer deletions that inactivate the potential to produce phage (our unpublished observation).

Table I gives some of the titers produced by wild-type (pMS2000) and mutant MS2 clones under various conditions. pMS2014 and pMS2043 contain a destabilized initiator hairpin. In pMS2023 and pMS2045 this hairpin is stabilized. Although we cannot account for all the observations, it is clear that pCOAT184 attenuates infection by our wild-type construct at 28°C by a factor of 10. Furthermore, we note that mutants pMS2045 and pMS2023, which make very little coat protein, have considerably reduced titers.

The general procedure for creating mutant phages and monitoring their evolution is described in Materials and methods. The generation of phages in $F^-$ strain M5219 is defined as cycle 1. These phages are then plated on $F^+$ strain KA797 (which does not contain pCOAT184) to separate the viable genotypes that emerged (cycle 2). A correlation between plaque size and genotype was not apparent. Usually ~10–20 plaques were selected for sequence analysis and for further growth in liquid medium. Occasionally, a plaque was dissolved in water and propagated by renewed plating. In mutants designed to have a stabilized

**Table I.** Titers of various MS2 cDNA constructs (p.f.u./ml)[a]

| | pMS2000[b] | pMS2000[b] +pCOAT184 | pMS2014 +pCOAT184 | pMS2023 +pCOAT184 | pMS2043 +pCOAT184 | pMS2045 +pCOAT184 |
|---|---|---|---|---|---|---|
| 28°C | $5 \times 10^{11}$ | $7 \times 10^{10}$ | $3 \times 10^{8}$ | $2 \times 10^{7}$ | $4 \times 10^{10}$ | $7 \times 10^{6}$ |
| 42°C | $2 \times 10^{11}$ | $2 \times 10^{11}$ | $3 \times 10^{8}$ | $1 \times 10^{8}$ | $1 \times 10^{11}$ | $9 \times 10^{6}$ |

[a]Appropriate dilutions of supernatants from cultures grown either overnight at 28°C or for 2 h at 42°C were tested on lawns of KA797 cells. Maximum values are shown here.

[b]Wild-type. Titers obtained for wild-type are similar to those described previously for cDNA clones of $Q_\beta$ (Taniguchi *et al.*, 1978; Shaklee *et al.*, 1988) and of MS2 (Shaklee, 1990).

helix, roughly half of the plaques obtained in cycle 2 were wild-type, presumably due to recombination with pCOAT184 at either the RNA (Palasingam and Shaklee, 1992) or the DNA level. These were discarded. In mutants carrying a destabilized initiator hairpin such wild-type plaques were seldom found. The evolution of the plaques containing mutant sequences is described below.

### *Evolution of stabilized mutants*

In the center of Figure 2 we present the structure of mutant 45.0. It contains five mutations relative to the wild-type (black boxes). Base-pair III is stabilized from A-U to G-C and pair XII from G·U to G-C, while the U·U mismatch at pair VI is replaced by C·A, a change which is presumably energetically neutral. The calculated ΔG° of this mutant hairpin is −10.6 kcal/mol, causing a drop in coat protein synthesis to 0.1% with respect to wild-type as measured previously in a partial MS2 cDNA clone (de Smit and van Duin, 1990). Figure 2 shows that there are at least six escape routes from this unfavorable sequence. In one evolutionary path (45.1) base-pair III is disrupted to A·C causing a ΔG° of −5.2 kcal/mol. Although this stability is predicted to

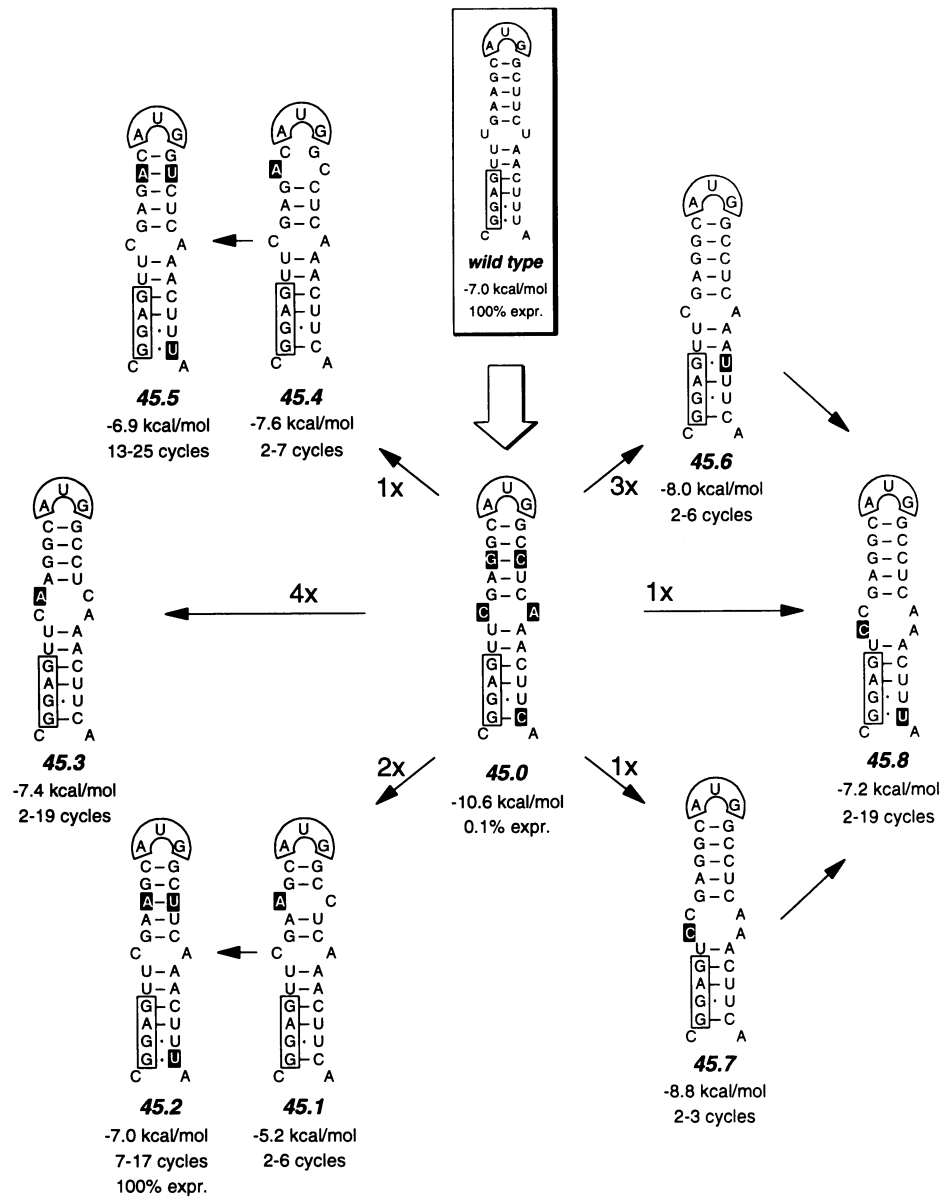


**Fig. 2.** Evolutionary pathways leading from non-viable mutant 45.0 to various viable species. Numbers at arrows pointing to revertants show how many plaques were found with this sequence. The wild-type structure is presented for convenience. Black boxes in the mutant sequence mark the differences from the wild-type. Black boxes in the revertant sequences mark the differences from the starting mutant. The indicated expression levels were measured in partial MS2 cDNA clones as described (de Smit and van Duin, 1990). Helix stabilities were calculated according to Freier *et al.* (1986) except when base-pair III was a mismatch. Instead of opening the loop to contain nine bases, we prefer to assume the structure is as drawn because ΔG° contributions for most loop sizes and loop sequences have not been determined experimentally. The calculation according to Freier *et al.* (1986) yields a ΔG° which is 0.6 kcal/mol lower than our alternative for these revertants. A value for the contribution of the G·U pair stacked upon the C·A terminal mismatch has not been published. Based on expression studies, de Smit (1994) has deduced that this value is −1.2 kcal/mol. [For a G-C pair stacked upon a C·A terminal mismatch a value of 2.0 kcal/mol was reported by Freier *et al.* (1986).] The difference between the helix stability used in this paper (−7.0 kcal/mol) and the value of −5.8 kcal/mol used in our previous work (de Smit and van Duin, 1990) is due to the fact that the present study was carried out at 37°C, whereas the previous one took place at 42°C. This temperature difference affects the ΔG° values. When more than one plaque with an identical sequence was found, the evolution of only one or two plaques was monitored.

produce wild-type quantities of coat protein, evolution does not yet stop. In the next four cycles two additional substitutions appear; the A·C at position III is changed further to A-U, while at the same time the bottom pair reverts to G·U. Both pairs are also present in the wild-type and as a consequence the stability is back to the wild-type value. The C·A mismatch in the middle remains untouched. This revertant, 45.2, was stable up to the end of the experiment, i.e. its sequence did not measurably evolve any further up to cycle 17. The yield of coat protein of revertant 45.2, as measured in partial MS2 cDNA constructs, was found to be the same as wild-type.

Revertant 45.7 shows that there is an alternative evolutionary path to regain the desired stability. First, pair VII turns into a C·A mismatch creating a symmetrical internal loop of four bases which reduces stability from −10.6 to −8.8 kcal/mol. This composition is not yet stable, presumably because it still produces too little coat protein. Indeed, two cycles later we find the bottom base-pair changed from G-C to G·U resulting in a stability of −7.2 kcal/mol. This revertant, 45.8, was stable for 19 cycles at which point the experiment was stopped.

In revertant 45.6 the stability strain on the initiator hairpin is eased by changing the wobble position in pair IX from G-C to G·U. This results in a ΔG° increase from −10.6 to −8.0 kcal/mol. This sequence appears stable for four cycles, but eventually evolves to that of 45.8 when propagated further. One plaque was found in which evolution to the stable revertant 45.8 had apparently been reached without resort to visible intermediates. This short cut requires two base changes in cycle 1, a not uncommon event (see 43.2).

A fifth pathway to viability is presented by revertant 45.3. Four plaques followed this route in which the fifth base-pair mutates from G-C to A·C. This relieves the stability strain in a way similar to that found above for revertants 45.7 and 45.8, which opened the seventh pair. A symmetrical internal loop containing four bases is created resulting in a calculated stability of −7.4 kcal/mol. This sequence appears stable as we could not detect any changes when this phage was propagated for 17 additional cycles.

Finally, a peculiar pathway is exhibited by revertants 45.4 and 45.5. Here, escape from the too stable helix is found by creating a mismatch at base-pair II leading to a stability of −7.6 kcal/mol. After having made this 'choice', there seems no easy way to get closer to wild-type stability, except by a new change in the second base-pair, i.e. from A·C to A-U (revertant 45.5, ΔG° = −6.9 kcal/mol). This revertant, in which the bottom pair has also changed to G·U, appeared to be stable up to cycle 25 when the experiment was stopped. This is the only time that we have found a revertant with a substitution that changes the coding properties (Ala2→Val).

It may be noted that none of the plaques (cycle 2) have retained the original mutant sequence. It would therefore appear that mutant 45.0 is not viable at all and that the existence of any progeny depends on base changes that have occurred in the first cycle in host M5219.

The second created mutant, 23.0, has a calculated stability of −8.7 kcal/mol (Figure 3) and it produces 3% of the wild-type amount of coat protein as measured using partial MS2 cDNA in an expression vector. The differences from wild-type are that pair III is broken and pair XII is further

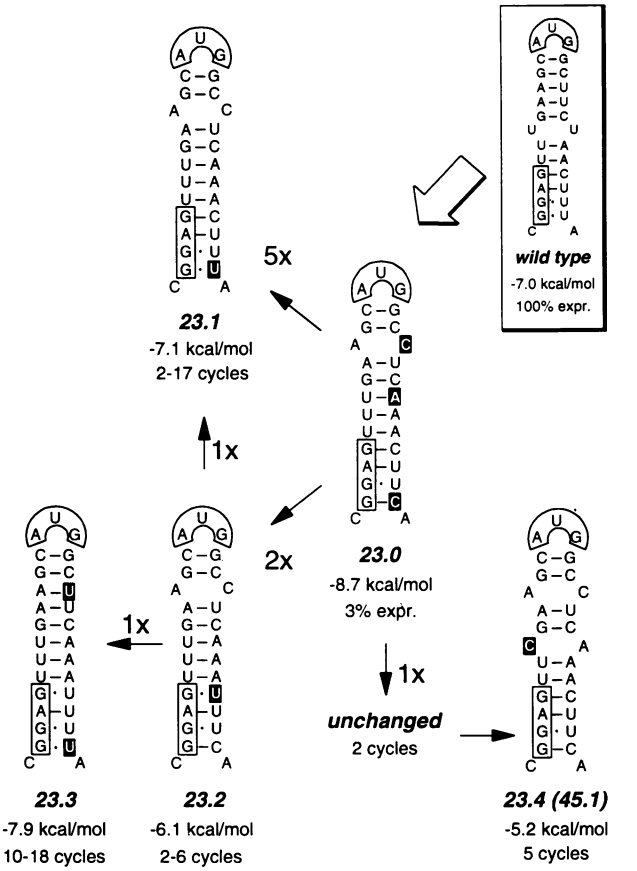


**Fig. 3.** Revertants of mutant 23.0. Revertant 23.4 is identical to 45.1. See legend to Figure 2 for further details.

stabilized while the middle U·U mismatch is changed into U·A, giving us the opportunity to assess the necessity of the internal loop *per se* at this specific position. Five plaques showed that there is an easy way in which this mutant can become a viable sequence: the bottom pair reverts to G·U which returns the stability to almost wild-type (−7.1 kcal/mol). The new revertant, 23.1, is stable up till 17 cycles, after which we stopped monitoring the sequence. The new sequence shows that there is no particular need for the mismatch to be at position VI. The phage is viable also with the mismatch at position III as long as the overall strength of the helix matches that of the wild-type.

An alternative pathway is shown in Figure 3 as revertant 23.2. Here, changing pair IX from G-C to G·U causes a reduction in stability to −6.1 kcal/mol. Although the sequence holds for several cycles, it still does not seem the best fitness that can be reached under the circumstances. Four cycles later two substitutions have occurred. Pair III becomes A-U and pair XII changes to the wild-type combination G·U. The end result is a new stability of −7.9 kcal/mol. Like its predecessor, revertant 23.2, this is still 0.9 kcal/mol off the wild-type value. Presumably, the latter species has an advantage over the former that is not related to the stability. Mutant 23.3 did not evolve any further for the next eight cycles (cycle 18). Note that in this sequence there is no mismatch in the hairpin. It is remarkable that changing base-pair VI from U·A to U·G would yield an apparently 'better' ΔG° of −7.2 kcal/mol. We do not know why this solution is not chosen.

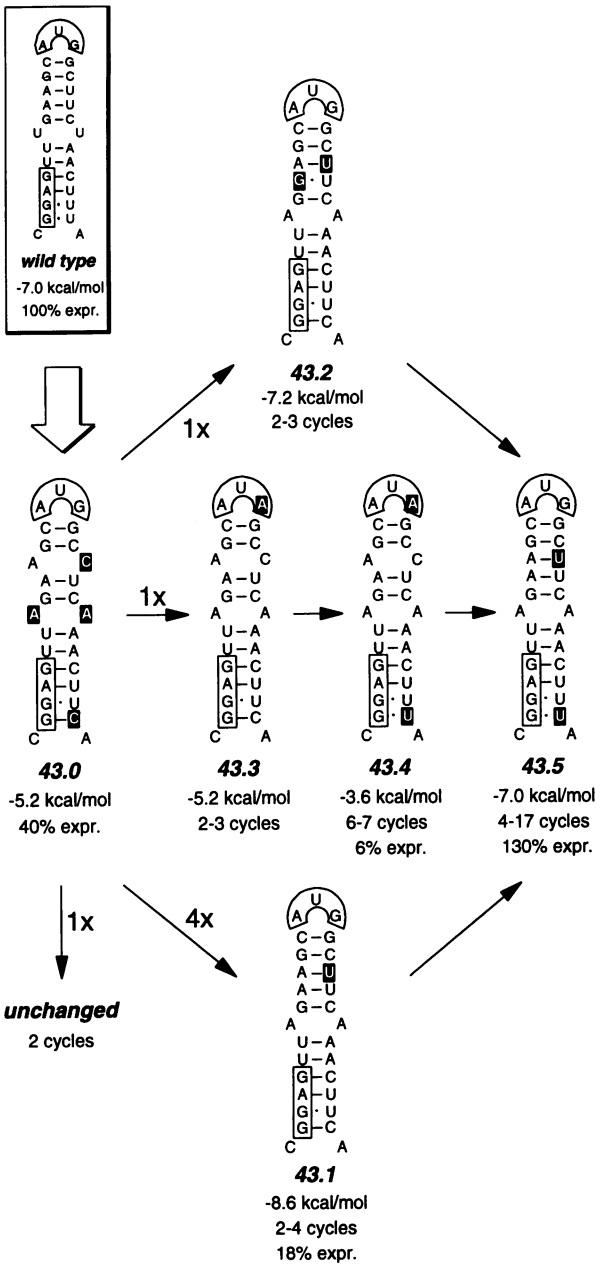


**Fig. 4.** Revertants of mutant 43.0. See legend to Figure 2 for further details.

One plaque was found containing the original mutant DNA sequence, showing that mutant 23 is not far off from a stable species. Eventually, the sequence gives way to the successful competitor 23.4. Here, base-pair VI is opened, changing the ΔG° in the favorable direction from −8.7 to −5.2 kcal/mol which should allow sufficient synthesis of the coat protein. This revertant (23.4) was lost, but it is identical to mutant 45.1 (Figure 2) and is thus predicted to change to the stable species 45.2.

### *Evolution of destabilized mutants*

As outlined in the Introduction, the simple rationale for the stabilized mutants to raise the ΔG° of the initiator hairpin to less negative values could be the need to produce sufficient coat protein. For the destabilized mutants there is *a priori* not an easily identifiable evolutionary pressure to change the ΔG°, but the fact that 45.1 evolved further to 45.2 (Figure 2) already indicated that a stability weaker than wild-type is unfavorable also. Below we show that this is a general pattern.

Mutant 43.0 (Figure 4) has base-pair III changed into the mismatch A·C, the U·U at position VI altered to A·A and the bottom pair stabilized to G-C, resulting in a ΔG° of −5.2 kcal/mol. For reasons not understood the expression study showed that this mutant had a coat protein yield of 40% rather than a value close to 100%. Plating the initial lysate of mutant 43.0 (cycle 1) revealed the presence of three new genotypes in seven plaques analyzed. Four plaques showed repair of the mismatch at position III to A-U (revertant 43.1). Thermodynamically speaking this is an overshoot to −8.6 kcal/mol. Two cycles of infection further, the wild-type stability was indeed realized by a G-C→G·U change at the bottom pair (revertant 43.5). The A·A pair in the middle position is left unaltered, in agreement with the general assumption that all mismatches destabilize a helix to the same degree (de Smit, 1994).

One plaque (revertant 43.2) revealed an alternative route to wild-type stability. It had sustained two changes, one at pair III, the other at pair IV which, in combination, raise helix strength to −7.2 kcal/mol. Within the next three rounds two more base substitutions occurred, converting revertant 43.2 into 43.5, an end situation also reached via the other pathway (Figure 4).

A rather peculiar development is represented by revertant 43.3. Here, the weak mutant helix is apparently compensated for by a change in the initiation codon to AUA. Subsequently, the hairpin is further destabilized at base-pair XII. Thereafter, base-pair III is closed and the AUG start codon restored (revertant 43.5). It is difficult to understand how such a pathway can exist, particularly in view of our finding that a partial MS2 cDNA clone that has the sequence of 43.4 produces only 6% coat protein relative to the wild-type analog. This pathway reveals another level of complexity that will be addressed in the future.

Finally, we have pursued the fate of mutant 14.0 (Figure 5) in which the third base-pair was changed from A-U to U·C. Return to the complete wild-type *sequence* would require two substitutions, whereas the wild-type *stability* can be reached simply if the U·C pair changes into U-A. This is what in fact happens in 15 out of the 18 plaques analyzed (Figure 5, revertant 14.1). This end situation can also be attained via the U·G intermediate at pair III, which already stabilizes the helix from −3.6 to −6.6 kcal/mol (revertant 14.2). As described above for the other mutants, here too more solutions exist to escape from the apparently unfavorable ΔG°; one other alternative is shown by revertant 14.4. Basically, the solution is that pair VI rather than pair III mutates to U-A. This yields a wild-type ΔG° and this mutant resembles 23.1, except that it has U·C instead of A·C at base-pair III. It is interesting that we have been able to isolate a metastable intermediate (revertant 14.3) where base-pair XI is changed from G·U to A-U. This change raises stability slightly in the desired direction but it decreases the SD complementarity from GGAG to GAAG, a change that turns a G-C match with the anti-SD sequence in 16S rRNA into an A·C mismatch. We have shown recently that a weaker SD interaction suffices when the helix containing the ribosome binding site is destabilized accordingly (de Smit and van Duin, 1994); this is possibly why revertant 14.3

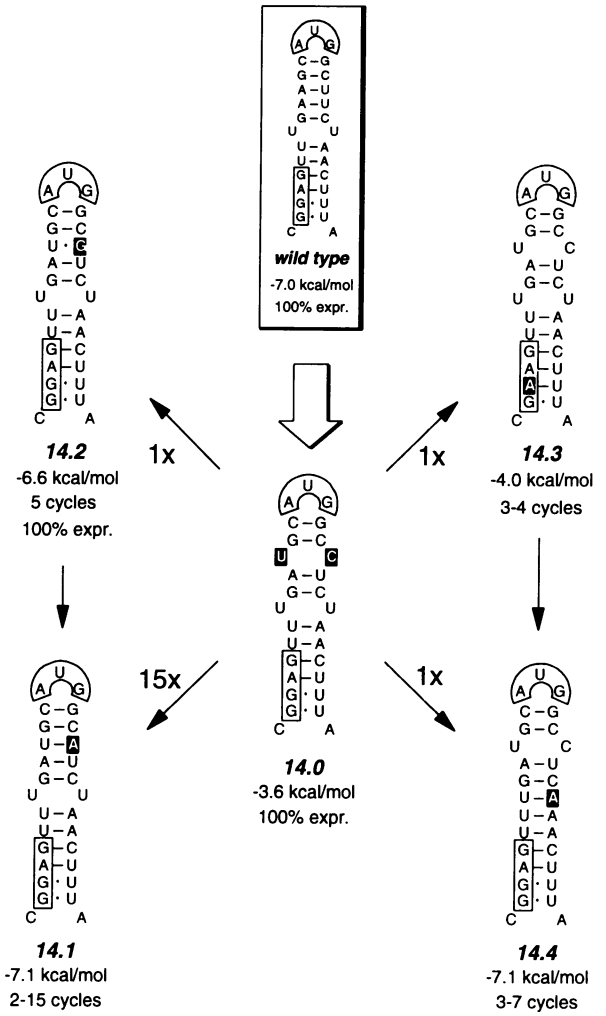


**Fig. 5.** Revertants of mutant 14.0. See legend to Figure 2 for further details.

is viable enough to exist transiently. On the other hand, we cannot exclude the possibility that revertant 14.3 is analogous to 43.3 and 43.4, where too weak a hairpin is initially compensated for by reducing translation. As shown in Figure 5, revertant 14.3 quickly evolved to the more stable sequence of 14.4.

### *Evolution of non-initiator hairpins*

To understand the nature of the selection pressure that causes the energetically compromised hairpin mutants to revert to wild-type $\Delta G°$, we also introduced stabilizing and destabilizing substitutions in the wobble positions of the second coat-gene hairpin (Figures 1 and 6). This hairpin is phylogenetically conserved in group A phages and was confirmed by chemical and enzymatic probing (Skripkin *et al.*, 1990). Earlier, we did not find any measurable effect of these mutations on coat protein synthesis in partial MS2 cDNA clones (M.H.de Smit and J.van Duin, unpublished results). Pursuing the evolution of these mutants may indicate whether or not a fixed hairpin stability is in itself a target for strong selection. Our results show that the destabilized mutant 30 returns to a mixed mutant/wild-type sequence after two cycles and is fully wild-type after five cycles (Figure 6). For the stabilized mutants 31 and 29 no changes were

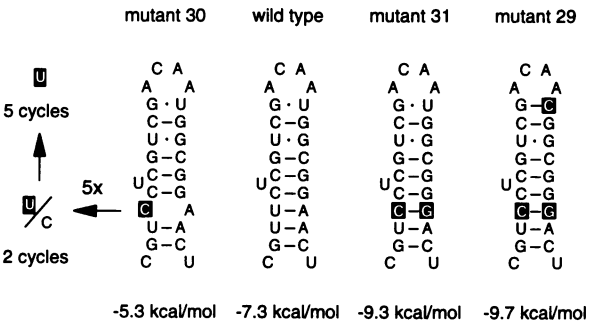


**Fig. 6.** Revertants of mutants in the second coat gene helix. Mutants 31 and 29 were unchanged after 13 cycles.

observed for 13 cycles. A similar experiment was performed for a phylogenetically proven hairpin in the replicase gene. Also here, (partial) compensation for destabilization was already noticed after a few cycles, whereas strengthened stem−loop structures took much longer to revert (data not shown). It seems that the pressure on low stability is a general and strong one, but selection against high stabilities is not very strong in non-initiator helices.

### *Competition between revertants and wild-type*

To get an impression of the fitness of our revertants we mixed wild-type MS2 with revertants 23.1, 14.1, 43.5 or 45.2 in a ratio of 1:200 based on plaque-forming units. Subsequently, each mixture was allowed to complete five cycles. As before, phage were isolated and the RNA sequenced with reverse transcriptase. In none of the preparations could we detect the presence of the wild-type sequence, although the degree of dilution and the number of cycles would have allowed the wild-type to take over visibly if it had had a sufficient advantage over the revertant (see below). The next experiment was more demanding on the mutants: they were mixed with wild-type in a 1:1 ratio. Now, revertants 23.1, 45.2 and 43.5 became extinct after three cycles, while 14.1 competed well. Its sequence was still clearly present in the final RNA preparation.

### *Related and unrelated base substitutions at distant sites in revertants*

In view of the low accuracy of RNA replication there are two questions that need some attention. One is whether any of the revertants had acquired additional phenotypical compensation as a result of substitutions outside the region under scrutiny here. Although it is not feasible to analyze the complete genome of all revertants, we have determined partial sequences of several revertants. The analysis involved a part of the M site in the replicase gene around nucleotide 2050 where protein S1 was cross-linked and where the replicase protein is supposed to bind, and also concerned sequences at the 5′ and 3′ ends (Boni and Isaeva, 1988; Skripkin *et al.*, 1990). No deviations from the wild-type sequence were found.

The other question concerns whether random base changes occur. Here, the answer is positive. One of the fifteen plaques representing revertant 14.1 contained an additional U→C base change at codon wobble position 1394. This revertant was stable for as long as we monitored the sequence (14 rounds). However, when placed in competition with the wild-type in a ratio of 200:1, five cycles sufficed to turn

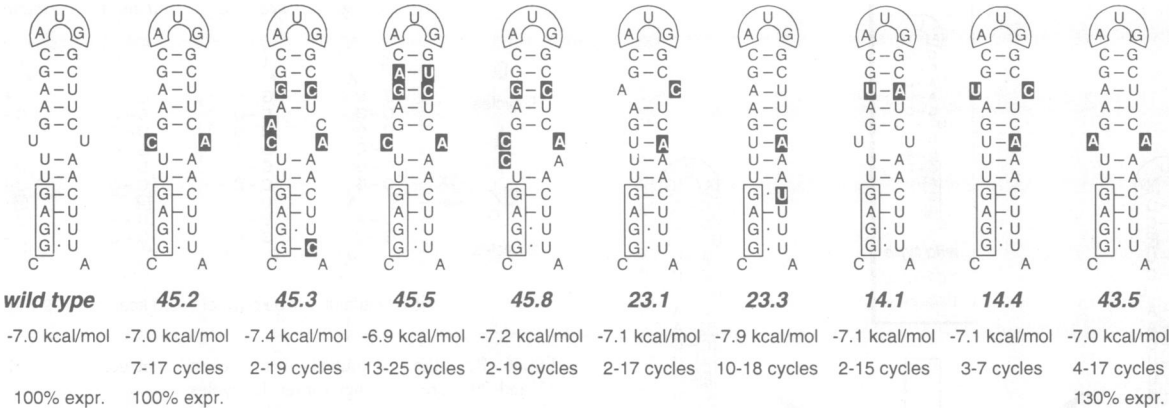


**Fig. 7.** Compilation of coat initiator helices in stable revertants.

it into a minority species in the mixture. This behavior contrasts with that of the regular 14.1 revertants which persisted in this competition experiment (see above). This result shows that the substitution is the result of a random error; it does not 'improve' the phenotype.

## Discussion

A previous expression study suggested that the stability of the initiator hairpin of the coat protein gene of phage MS2 is carefully tuned. Any mutation that increased its stability led to a dramatic loss in protein yield whereas substitutions that weakened the structure had no effect. The hairpin thus seemed endowed with the maximal stability that still permits maximal ribosome loading. That analysis, carried out on a partial MS2 cDNA clone, revealed that translation did not respond to the identity of the bases that we substituted but to the $\Delta G°$ that was the result of these substitutions (de Smit and van Duin, 1990). The present analysis complements and extends that study. Using an infectious MS2 cDNA clone we have now introduced stabilizing and destabilizing mutations into the initiator hairpin in the live virus and followed the Darwinian evolution of the new mutants to a viable species. Because some of the starting mutants would only synthesize minimal amounts of coat protein (45.0 and 23.0), a helper plasmid supplying coat protein *in trans* was employed to provide the phage with some starting capital. After this first cycle, which was carried out in a non-permissive $F^-$ host, the phages formed were tested by plating them on an $F^+$ host that did not provide extra coat protein. Since RNA-dependent RNA polymerases have a low copying fidelity ($\sim 10^{-4}$ per nucleotide) new sequences arose even in the first cycle and, if viable, manifested themselves as plaques after plating. Survivors were then grown further in liquid culture to monitor their evolution to a stable species, i.e. a species whose sequence did not change any further, at least not for the duration of the experiment, which was $\sim 19$ infection cycles.

We observed that both stabilized and destabilized initiator hairpins were strongly selected against. All reverted to a stability close to or the same as that of the wild-type (see Figure 7 for a compilation of revertant structures). The end situation was achieved in a single step or through a number of intermediates, characterized by a helix $\Delta G°$ closer to wild-type than the initial mutant sequence. Recovery of wild-type stability was achieved by secondary mutations that compensated for the sustained loss or gain in helix stability. In parallel with these changes the vitality of the revertants increased, as witnessed by a rise in phage titer. By the time the sequence had reached its new equilibrium, the titer equalled that of the wild-type (data not shown).

With three exceptions (revertants 45.5, 43.3 and 14.3), substitutions occured in the stem at codon wobble positions and outside the SD sequence. Assuming that the final revertants produced the same amount of coat protein as wild-type (for some this was measured and found to be true), one conclusion of the present study is that, except for the start codon and the SD sequence, the ribosome appears blind to the sequence, but very sensitive to helix stability. A second conclusion derives from the finding that substitutions in wobble positions are encountered. The wild-type early coat protein codons GCU, UCU and AAC are preferentially used in highly expressed genes in *Escherichia coli*, whereas their synonyms GCC, UCA and AAU are avoided and occur with up to 20 times reduced frequency in such genes (Andersson and Kurland, 1990). The fact that we found these less preferred codons in many revertants indicates that the influence of codon usage on ribosomal clearance of the initiation region is, at least for the codons changed here, marginal relative to the structure effects.

The prevalence of secondary over primary structure is also consistent with our finding that all of the viable revertants arose by compensatory substitutions in the stem region. A ribosome initiating at the coat gene start protects 12 nt downstream and 20 nt upstream of the AUG codon against RNase attack (Figure 1). This includes the presumably single-stranded region of 7 nt upstream of the initiator hairpin. If base-specific contacts were to exist one would expect to find compensatory mutations in this single-stranded region too, as these could further tune initiation rates. This RNA region is non-coding and thus free to evolve. We have not found any mutations here, suggesting that base-specific interactions do not exist outside the SD region and the AUG codon.

As well as indicating which regions are free to evolve, the obtained revertants also show us what constraints exist. For instance, one can envisage many helices with wild-type stability if different amino acids are allowed at the N-terminus of the coat protein. Only one of these possibilities was found among our revertants (45.5), indicating that the nucleic acid sequence is severely constrained in this part of the RNA by the protein although the N-terminus of the coat

protein is not implicated in dimer or capsid formation or in replicase repression (Valegård *et al.*, 1990; Peabody, 1993).

Our finding that the revertants were biologically stable would suggest that such helix variants could be found in nature among MS2 relatives. Surprisingly, this is not true. The group I phages MS2, f2, R17, M12 and JP501 have been isolated independently in different parts of the world and at different times. Yet MS2, M12 and JP501 have an identical sequence in the region under consideration here, whereas the remaining two have only one substitution each; f2 has U·C at base-pair VI and in phage R17 base-pair X is G·U instead of A-U. Presumably, the selective advantage of the wild-type MS2 over the revertants is so small that our evolutionary experiment did not last long enough to reveal the difference. Such a view is supported by the competition experiments. When mixed in equal ratios, the wild-type quickly displaced all of the revertants (except one) showing its selective advantage. However, when MS2 and revertant are mixed in a 1:200 ratio, the wild-type does not outgrow the revertant within five cycles, showing that the advantage is small. In this respect it should be realized that there is an additional reason why the revertants are stable on their own. Not only do they differ from the wild-type in at least two positions, but, except for 45.2, every single substitution towards the wild-type sequence results in a large change in $\Delta G°$ away from the desired optimum. Thus, revertants are trapped in an evolutionary environment that prevents their escape to a 'better' sequence.

The present analysis also provides independent support for the existence of the initiator hairpin since our revertants show base changes that maintain the existence of the stem−loop structure. Thus, the procedure described here, in which the distortion in an RNA genome is repaired, can also be used to demonstrate the presence of base-paired regions whose existence cannot be proven otherwise. It would suffice to introduce mutations in the region of interest and analyze if the disturbance is compensated for in the presumed complementary region. This 'instant phylogeny' can be particularly useful for sequences that show little or no variation in nature and has been used as such in a few instances (Macadam *et al.*, 1992; Tsai and Dreher, 1992).

The present study also clearly illustrates one of the problems in applying the phylogenetic approach to deduce secondary structure features in phage RNA. As selection is targeted at stability rather than at the precise shape of the helical segment, the exact position of mismatches in a helix is flexible as long as the optimal $\Delta G°$ is attained (for example, compare revertant 14.1 with 14.4, or 45.3 with 45.8). As a consequence it may not be possible to prove the existence of the initiator helix by the conventional criterion of two covariations per helix (James *et al.*, 1990). We have noted previously that stability appears preserved also for other helices in phage RNA, albeit less strict. This suggests that the criteria for proof by comparative analysis need to be extended to include thermodynamic parameters.

One important question is that of the selection pressure operating on the $\Delta G°$ of the initiator hairpin. Above we have discussed that hairpins that are too stable do not synthesize maximal amounts of coat protein and this seems a plausible enough reason for counter-selection. Nevertheless, we have carried out a control experiment in which the evolution of a mutationally stabilized non-initiator hairpin was recorded. Reversion to wild-type $\Delta G°$ was not reached within 13 cycles, even though this would require only one or two transitions (mutants 29 and 31, respectively). Although we have little doubt that the wild-type helix confers a 'better' phenotype, the advantage may be so small that many more cycles would be required to reveal the difference.

The pressure that forces weakened initiator helices to raise their stability is presently not understood, but a few comments can be made. The rapid evolution of mutant 30 containing a non-initiator hairpin suggests, at least when compared with the stabilized analogs (mutants 29 and 31), that there is a general strong pressure against weak helices. This may derive from the need to separate mother and daughter strands during replication (Axelrod *et al.*, 1991) or to provide protection against RNases.

Another realistic possibility is that helices that are too weak distort the structure of the viral RNA by provoking alternative base-pairings with an unfavorable replication phenotype (Biebricher and Luce, 1992). Here, we would like to mention that the pattern of aspecific stops visible in primer extension experiments in some of the mutants containing destabilized initiator helices differs from those found in the majority of revertants and in the wild-type. This is a hint that the structure has changed.

A more speculative explanation for the reversion of the destabilized initiator hairpin may be found in the complex system that the RNA phages have developed to save their genome from the hazard of being translated and replicated at the same time. This problem is solved by competition between replicase and ribosome for a common internal site on the viral RNA, allowing only one of the two components to bind. In $Q_\beta$ RNA this question has been studied *in vitro* and the common site, the S site, overlaps the start region of the $Q_\beta$ coat protein gene (Meyer *et al.*, 1981). For the group I phages such studies have not been carried out, but by analogy one could suppose that the MS2 replicase would bind the coat initiator region to provide the necessary competition. One could imagine that MS2 replicase binds to the coat start better when the stem is weaker. This would upset the delicate binding balance between replicase and ribosome, possibly leading to decreased virus production. A slightly different version of this model has been given elsewhere (de Smit and van Duin, 1993).

## Materials and methods

### *Bacterial strains*

Mutant and wild-type infectious MS2 cDNA clones were grown in M5219 (M72 $trpA_{am}$, $lacZ_{am}$, $Sm^r/\lambda dbio_{252}$, $cI_{857}\Delta H_1$), encoding the thermosensitive repressor ($cI_{857}$) and the transcriptional antitermination factor N (Remaut *et al.*, 1981). *E.coli* F$^+$ KA797 (*F'lacI$^Q$, pro/ara, Δlac-pro, thi*) was used as host for wild-type and mutant MS2 phages. All strains were grown on LC broth containing per liter 10 g bactotryptone, 5 g yeast extract, 8 g NaCl, 2 g $MgSO_4$, 140 mg thymine and 1 ml 1 M Tris−HCl, pH 7.6.

### *Plasmids*

pCOAT184 is derived from pACYC184 by cloning the MS2 coat protein gene (1221−1736) into the *Bam*HI site of the tetracycline resistance gene as described by Berkhout (1986). In this construct the coat protein is under control of the constitutive *tet* promoter which ensures a low level of coat protein sufficient to repress replicase gene translation by ~90%.

The construction of the full-length infectious MS2 cDNA clone, called pMS2000, will be described elsewhere (R.C.L. Olsthoorn). The 5′ end of MS2 cDNA in these clones is preceded by a G-tail and the 3′ end continues in an A-tail. As a negative control we prepared a clone, pMS2001, in which the internal *Sac*I fragment (1491−3387) was present in the reverse

orientation. Both constructs are under control of the $p_L$ promoter of phage λ. pMS2001 did not produce any phage.

Mutants of the infectious clone were constructed by replacing the *XbaI*(1303)–*BstXI*(1551) fragment of pMS2000 with the corresponding fragment from a series of mutant sequences used for coat protein expression studies (de Smit and van Duin, 1990). The presence of the mutations in the final constructs was verified by sequencing both DNA strands using the T7 sequencing kit of Pharmacia and a primer complementary to nucleotides 1409–1422 of the coat gene. Four mutant plasmids were constructed: pMS2014, 2023, 2043 and 2045. The last two digits are used to identify the corresponding phage mutants in the Results section as 14.0, 23.0, etc. and their evolutionary offspring are called revertants and are indicated as 14.1, 14.2, etc.

### *Sequence analysis of wild-type and mutant MS2 phages*

The phages present in the supernatant of a 1 ml culture were precipitated with 330 μl of 40% polyethylene glycol in 2 M NaCl. The pellet was dissolved in 200 μl of TE (10 mM Tris–HCl, pH 7.6 and 0.1 mM EDTA) and extracted with 300 μl of phenol/chloroform (2:1). The RNA was precipitated with 2.5 vols of ethanol and dissolved in 20 μl of distilled water.

1–2 μl of this solution was used for each sequence reaction, which was carried out by primer extension with reverse transcriptase (Promega) essentially as described by Skripkin *et al.* (1990). An oligonucleotide complementary to nucleotides 1409–1422 in the coat gene was used as primer and enabled screening of ~300 nt on the MS2 genome for differences with the wild-type.

### *Measuring phage evolution*

Routinely, the procedure to produce mutant phages involves growing *E.coli* strain M5219 containing a plasmid with mutant MS2 cDNA and pCOAT184 at 28°C until the $OD_{650}$ is 0.2. (In this host reinfection cannot take place because of the absence of F pili.) Then the culture is shifted to 42°C. After 2 h it is centrifuged to remove intact cells and debris; the supernatant containing the phages is used for two purposes. One part is precipitated and the pellet used to determine the RNA sequence. Another fraction of the supernatant is plated on KA797 ($F^+$) to separate the genotypes that may have developed during growth in M5219. Individual plaques are dissolved in $H_2O$ and grown further on plates or in culture for the indicated number of cycles. We define growth in M5219 as cycle 1, and the first growth as plaques in KA797 as cycle 2. (The amount of phages present in a plaque is too small to be sequenced. Accordingly, they had to be amplified by growth in liquid culture. It is not known whether additional substitutions arise during this extra cycle.) Each subsequent cycle started with the inoculation of 3 ml KA797 ($OD_{650}$ = 0.07; $10^7$ cells) with $10^3$–$10^5$ p.f.u. and was ended by overnight growth at 37°C in a test tube or on a plate. Alternatively, a plaque was propagated further on a bacterial lawn in solid medium. Occasionally, the amount of phage used to seed the next cycle was increased to $10^7$ p.f.u. when we expected this to increase our chances of picking up double or triple mutants. As a consequence, the term 'cycle' is loosely defined in this study and cannot be related to the number of phage generations in a simple way. Since the reversion mutations arise by chance we have not attempted to standardize our procedures rigorously. Sometimes an overnight culture of M5219 at 28°C was used as cycle 1. No differences in titer from phage creation at 42°C were observed.

## Acknowledgements

We thank Drs Pleij and de Smit for comments on this manuscript. Phage M12 was a generous gift from Dr R.Konings and mutant MS2 cDNA was supplied by Dr de Smit. We are grateful to Drs A.Hirashima and K.Furuse for making available to us phage JP501 from the Watanabe collection. Steven Zoog is acknowledged for help in some of the experiments. N.Licis was supported by an EC grant in the Go West PECO programme (ERB3510PL922400).

## References

Andersson,S.G.E. and Kurland,C.G. (1990) *Microbiol. Rev.*, **54**, 198–210.
Axelrod,V.D., Brown,E., Priano,C. and Mills,D.R. (1991) *Virology*, **184**, 595–608.
Barrera,I., Schuppli,D., Sogo,J.M. and Weber,H. (1993) *J. Mol. Biol.*, **232**, 512–521.
Berkhout,B. (1986) Ph.D. Thesis, Leiden University.
Biebricher,C.K. and Luce,R. (1992) *EMBO J.*, **11**, 5129–5135.
Boni,I.V. and Isaeva,D.M. (1988) *Dokl. Akad. Nauk SSSR*, **298**, 1015–1018.
de Smit,M.H. (1994) Ph.D. Thesis, Leiden University.
de Smit,M.H. and van Duin,J. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 7668–7672.
de Smit,M.H. and van Duin,J. (1993) *Mol. Microbiol.*, **9**, 1079–1088.
de Smit,M.H. and van Duin,J. (1994) *J. Mol. Biol.*, **235**, 173–184.
Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 9373–9377.
Guisez,Y., Robbens,J., Remaut,E. and Fiers,W. (1993) *J. Theor. Biol.*, **162**, 243–252.
James,B.D., Olsen,G.J. and Pace,N.R. (1989) *Methods Enzymol.*, **180**, 227–239.
Macadam,A.J., Ferguson,G., Burlison,J., Stone,D., Skuce,R., Almond,J.W. and Minor,P.D. (1992) *Virology*, **189**, 415–422.
Meyer,F., Weber,H. and Weissmann,C. (1981) *J. Mol. Biol.*, **153**, 631–660.
Mills,D.R., Dobkin,C. and Kramer,F.R. (1978) *Cell*, **15**, 541–550.
Palansingam,P. and Shaklee,P.N. (1992) *J. Virol.*, **66**, 2435–2442.
Peabody,D.S. (1993) *EMBO J.*, **12**, 595–600.
Priano,C., Kramer,F.R. and Mills,D.R. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 321–330.
Remaut,E., Stanssens,P. and Fiers,W. (1981) *Gene*, **15**, 81–93.
Remaut,E., De Waele,P., Marmenout,A., Stanssens,P. and Fiers,W. (1982) *EMBO J.*, **1**, 205–209.
Shaklee,P.N. (1990) *Virology*, **178**, 340–343.
Shaklee,P.N., Miglietta,J.J., Palmenberg,A.C. and Kaesberg,P. (1988) *Virology*, **163**, 209–213.
Skripkin,E.A. and Jacobson,A.B. (1993) *J. Mol. Biol.*, **233**, 245–260.
Skripkin,E.A., Adhin,M.R., de Smit,M.H. and van Duin,J. (1990) *J. Mol. Biol.*, **211**, 447–463.
Steitz,J.A. (1969) *Nature*, **224**, 957–964.
Taniguchi,T., Palmieri,M. and Weissman,C. (1978) *Nature*, **274**, 223–228.
Tsai,C.-H. and Dreher,T.W. (1992) *J. Virol.*, **66**, 5190–5199.
Valegård,K., Liljas,L., Fridborg,K. and Unge,T. (1990) *Nature*, **345**, 36–41.
van Duin,J. (1988) In R.Calender (ed.), *The Bacteriophages*. Plenum Press, New York, Vol. 1, pp. 117–167.
van Himbergen,J., van Geffen,B. and van Duin,J. (1993) *Nucleic Acids Res.*, **21**, 1713–1717.
Varenne,S., Buc,J., Lloubès,R. and Lazdunski,A. (1984) *J. Mol. Biol.*, **180**, 549–576.
Witherell,G.W., Gott,J.M. and Uhlenbeck,O.C. (1991) *Progr. Nucleic Acid Res. Mol. Biol.*, **40**, 185–220.

*Received on January 14, 1994; revised on March 11, 1994*