EVOLUTION AND ESSENTIAL FEATURES OF mRNA TRANSLATION INITIATION REGIONS IN PROKARYOTES. AN ANALYSIS OF PHAGE MS2 RNA

NORMUNDS LICIS

Biomedical Research and Study Centre

UNIVERSITY OF LATVIA
RIGA 1998

UNIVERSITY OF LATVIA

Normunds Licis

EVOLUTION AND ESSENTIAL FEATURES OF mRNA TRANSLATION INITIATION REGIONS IN PROKARYOTES. AN ANALYSIS OF PHAGE MS2 RNA

Author: .....**Normunds Licis**

Promoter: ..... **Valdis Berzins**
Professor, Dr. habil. biol.

Biomedical Research and Study Centre

RIGA 1998

University of Latvia

Habilitation and Promotion Council

Biology II

Riga, Latvia

Chairman: ..... **Elmars Grens**
Professor, Dr. habil. biol.

Opponents: **Mart Ustav** (Estonian Biocentre, Tartu)
Professor, Dr. habil. biol.

Kestutis Sasnauskas (Institute of Biotechnology, Vilnius)
Professor, Dr. habil. biol.

Pauls Pumpens (Biomedical Research and Study Centre,
Professor, Dr. habil. biol University of Latvia, Riga).

98-5858

Dissertation is accepted for the commencement of the degree of
Dr. biol. (in *Molecular Biology*)

Commencement: **January**.....28....., **1998**

KOPSAVILKUMS

Prokariotu mRNS translācijas iniciācijas rajonu evolūcija un būtiskās pazīmes. Fāga MS2 RNS analīze.

Normunds Līcis

Šī promocijas darba forma ir publicēto vai publicēšanai sagatavoto zinātnisko rakstu sērija. Darbs ir izstrādāts Latvijas Universitātes Biomedicīnas pētījumu un studiju centrā laika posmā no 1991. līdz 1997. gadam. Daļa eksperimentu veikta Leidenes Universitātes Gorlaeus Laboratorijā, Nīderlande.

Gēnu translācijas precizitāte un efektivitāte prokariotos ir atkarīga no sākotnējā translācijas posma, kura laikā ribosomas saistās ar informācijas RNS un atpazīst uz tās translācijas iniciācijas rajonu (TIR). Translācijas iniciācijas efektivitāti savukārt nosaka TIR īpašības un to raksturošana ir nozīmīga ne vien teorētiskajā, bet arī praktiskajā plāksnē, kā piemēram pielietojumam biotehnoloģijā. Mūsu pētījums ir veltīts RNS TIR būtisko pazīmju noskaidrošanai un analīzei, izmantojot bakteriofāgu MS2 un fr vienpavediena RNS. Šo baktēriju vīrusu genomu veido minētā RNS molekula un tās gēnu ekspresija tiek kontrolēta galvenokārt translācijas līmenī.

Darbā tika pielietotas divas atšķirīgas eksperimentālās pieejas. Pirmā tēžu daļa ir veltīta *in vitro* analīzēm, bet otrajā daļā aprakstītajos pētījumos mēs izmantojām vairumam RNS vīrusu raksturīgo īpašību – augsto adaptācijas potenciālu – lai risinātu aplūkojamās problēmas ar dabiskās evolūcijas procesa palīdzību.

In vitro dati ļauj mums secināt, ka bakteriofāgu MS2 un fr RNS translācijas iniciācijai būtiskā informācija ir lokalizēta ne vairāk kā 16 nukleotīdu garā fragmentā, kas satur translācijas iniciācijas kodonu AUG un SD-sekvenci. Evolūcijas pētījumu rezultāti, kas iegūti, analizējot fāga MS2 apvalka proteīna un replikāzes gēna mutaģenizētus TIR, norāda, ka izņemot SD-rajonu un translācijas iniciācijas kodonu, citas nukleotīdu sekvences īpašības ir samērā nebūtiskas translācijas iniciācijas procesam. Savukārt nozīmīga TIR efektivitāti ietekmējoša pazīme ir tā RNS otrējās struktūras stabilitāte. Dotais darbs pierāda arī agrāk izteiktos pieņēmumus, ka bakteriofāga RNS-RNS tālās distances mijiedarbības ir svarīgas replikāzes gēna translācijas kontrolei. Bez tam eksperimentālie dati liecina, ka, kaut arī SD-secība ir būtiska translācijas iniciācijai, tā nav absolūti nepieciešama funkcionālai ribosomu saistībai.

Darba rezultāti ir apkopoti 5 rakstos, 3 no kuriem ir publicēti, bet 2 sagatavoti publicēšanai starptautiski recenzējamos žurnālos, kā arī ziņoti 6 starptautiskos kongresos un konferencēs.

SUMMARY

Evolution and essential features of mRNA translation initiation regions in prokaryotes. An analysis of phage MS2 RNA.

Normunds Līcis

The form of the thesis is the summary of published or prepared for publication scientific articles. The work was carried out in Biomedical Research and Study Centre of Latvian University in the period between 1991 and 1997 years and some of the experiments were done in Gorlaeus Laboratories of Leiden University, The Netherlands.

Gene expression in prokaryotes ultimately depends on the efficiency and the accuracy of translation that in turn is determined at the stage when ribosomes recognize and bind to mRNA translation initiation region (TIR). The properties of TIR are the major determinant of the efficiency of translation initiation and their revealing is of importance both from theoretical and practical viewpoint, for example application in biotechnology. The present work is devoted to both aspects of general relevance for control of prokaryotic translation by features of mRNA TIR and particular molecular basis of translational control in the subjects of this study – single stranded RNA bacteriophages MS2 and fr. Phage MS2 is a member of a group of bacterial viruses in which the genetic information is stored in and expressed directly from RNA genome and thus the gene expression in these organisms is controlled mainly at the translational level. Two general approaches have been used in the present work. The first part of the thesis concerns with *in vitro* experiments whereas in the second part we have exploited a characteristic feature of most RNA viruses – a high adaptability - to test the phage RNA properties essential for translational control by natural evolution. *In vitro* data allowed us to conclude that all the important information for an efficient initiation of the replicase gene translation in phage MS2 and its close relative phage fr TIR does not expand outside a short RNA fragment in length of 16 nucleotides. The results also imply that within this minimal active initiation region only the translation initiation codon AUG and a 5'-region containing the so-called SD-sequence are important for the efficiency of the initiation process. The results of evolutionary studies on phage MS2 coat protein gene and replicase protein gene TIRs show that apart SD-sequence and initiation codon there are no other sequence-specific features of a general relevance for the efficiency of translation initiation. At the same time an essential property of an initiation site was found to be the strength of secondary structure that occludes TIR and may exert strong negative effect on the initiation rate. It was also shown, in agreement with earlier suggestions, that the global phage RNA folding is important for the control of the replicase gene translation. We observed as well that the extent of suppression is strongly dependent on sequence/structure of the region involving the replicase gene start or its surroundings. Our data also show that although SD-sequence is essential for the translation initiation it is not an absolute requirement for ribosome binding to take place.

The results of the this work are compiled in 5 articles, 3 of which have been published but 2 are prepared for publication in ISI journals, and also presented in 6 international congresses and conferences.

Zusammenfassung

Evolution der Translations- Initiationsregion prokaryontischer mRNA und ihre wesentlichen Symptone. Analyse der Phagen MS2 RNA.

Normunds Licis

Die Promotionsarbeit umfaßt veröffentliche und für die Veröffentlichung vorbereitete wissenschaftliche Betrage. Die Arbeit wurde im Biomedizinischen Forschungs- und Studienzentrum der Universität Lettland und teilweise auch im Laboratorium von Gorlaeus der Universität Leiden von 1991 bis 1997 durchgeführt.

Die Genauigkeit und Effektivität der Gentranslation in Prokaryoten hängt vom Beginn der Translation ab.Wenn die Ribosomen sich mit der Messenger- RNA binden und auf ihr die Translations- Initiationsregion (TIR) erkennen. Die Translations-Initiations Effektivität ihrerseits bestimmt die Eigenschaften der TIR, deren Charakterisierung nicht nur von theoretischer sondern auch praktischer Wichtigkeit, z.B., in der Biotechnologie ist. Gegenstand unserer Forschung sind die wessentlichen Eigenschaften und Analyse der RNA TIR, wobei einsträngige RNA der Bakteriophagen MS2 und fr verwendet wurden. Das Genom dieser Viren besteht aus dem Molekül der erwähnten RNA, und die Expression der Gene wird hauptsächlich auf dem Translationsniveau kontrolliert. Wir verwendeten zwei verschiedene experimentelle Methoden. Der erste Teil der Doktorarbeit ist der *in vitro* Analyse gewidmet. Im zweiten Teil der beschriebenen Experimente benutzen wir die der Mehrzalt von Viren charakteristischen Eigenschaft - das hohe Adaptionspotential um unser Problem mit Hilfte des natürlichen Evolutionsprozesses zu lösen.

Die *in vitro* Resultate führen zur Schlußfolgerung, daß die für die Translation- Initiation der RNA der Bakteriophagen MS2 und fr wesentliche Information sich in einem 16 Nucleotide nicht überschreitenden Fragment lokalisieren läßt. Dieses Fragment enthält das Translations- Initiationscodon AUG und die SD-Sequenz.

Resultate der Evolutionsforschung, die durch die Analyse des Hüll- Protein des Phages MS2 und mutagener TIR des Replikasen Genes erhalten wurden, zeigen, daß mit Ausnahmen des SD- Bereiches und des Translations- Initiations Codones, andere Eigenschaften der Nucleotidsequenzen verhalnismäßig unwichtig im Translation- Initiation Prozess sind. Anderseits ist die Stabilitat der RNA Sekundärstruktur von Wichtigkeit, da sie großen Einfluß auf die TIR Effektivität hat Wir konnten beweisen, daß die existierende RNA- RNA gegenseitige Einwirkung über weite Entfernungen für die Kontrolle der Translation des Replikasengenes von Bedeutung ist. Auserdem konnte gezeigt werden, daß die SD- Sequenz, obwohl wichtig bei der Initiation der Translation, nicht absolut nötig das für Funtionieren der Ribosomen- Bindung ist.

Die Forschungsresultate sind in 5 Publikationen zusammengefaßt, von denen 3 schon veröffentlicht und 2 zur Publikation in international anerkannten Zeitschriften eingereicht sind. Es wurden 6 Vorträge auf internationalen Kongressen und Konferenzen gehalten.

LIST OF ORIGINAL PAPERS

This thesis is based on the following papers, which are referred to by their Roman numerals:

I Stankevich, E., Kumpins, V., **Licis, N.**, Klovins, J. & Berzins, V. (1995) Template activity of oligoribonucleotides sythesized by the phosphoramidite method using 2'-O-tetrahydropyranyl protecting groups. *Nucleosides & nucleotides* **14**, 1047-1048.

II Kumpins, V., **Licis, N.**, Renhof, R., Stankevich, E., & Berzins, V. (1995) Synthetic 15-19- meric oligoribonucleotides as models for translation initiation of the phage fr replicase gene. *Bioorg. Chem.* **21**, 920-924 (in Russian).

III Olsthoorn, R.C.L., **Licis, N.** & van Duin, J. (1994) Leeway and constraints in the forced evolution of a regulatory RNA helix. *The EMBO Journal* **13**, 2660-2668.

IV **Licis, N.**, Balklava, Z. & Berzins, V. (1998) Evolution of translation initiation region of the phage MS2 RNA replicase gene from scratch. *Molecular Microbiology* (prepared for publication).

V **Licis, N.**, Balklava, Z. & Berzins, V. (1998) Long range translational coupling in single stranded RNA bacteriophages: an evolutionary analysis. *Nucleic Acids Research* (submitted).

LIST OF ABBREVIATIONS

A_{670}	absorbance at 670 nm
AMV	avian mieloma virus
bp	basepair(s)
BSA	bovine serum albumin
cDNA	copy DNA
dNTP	deoxynucleosidetriphosphates
ds	double-strand
E.coli	Escherichia coli
EF	elongation factor
fMet-tRNA	N-formylmethionyl-tRNA
IF	translation initiation factor
mRNA	messenger RNA
N	any nucleotide
nt(s)	nucleotide(s)
NTP	nucleoside triphosphates
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
R	purine
RBS	ribosome binding site
RlacZ	replicase/β-galactosidase gene fusion
rRNA	ribosomal RNA
RNase	ribonuclease
RT	reverse transcriptase
S1	ribosomal protein S1
S_1	nuclease S_1
SD	Shine-Dalgarno sequence
TIR	translation initiation region
T7	bacteriophage T7
Tris	tris(hydroxymethyl)-aminomethane
tRNA	transfer nucleic acid
16S r RNA	ribosomal RNA of small 30S ribosomal subunit of E.coli
wt	wild type

The choice is easy for the phage
but difficult for us.

Jan van Duin,
personal communication.

CONTENTS

3. Leeway and constraints in the forced evolution of a regulatory RNA helix (1994). *The EMBO J.* **13**, 2660-2668.
4. Evolution of a translation initiation region from scratch (1998). Prepared for publication.
5. Long-range translational coupling in single stranded RNA bacteriophages: an evolutionary analysis (1998) *Nucl. Acids Res.*, submitted.

INTRODUCTION

The success in maintenance and propagation of life depends on appropriate gene expression. Various proteins are required in entirely different quantities, often only at distinct periods or under certain conditions. To attain this flow of the genetic information from DNA to RNA (transcription) and from RNA to proteins (translation) must be somehow regulated. At least in prokaryotes, which mRNA is usually polycistronic *i.e.* caries reading frames for several proteins, translation, and especially its initial stage, is of utmost importance for the control. Individual prokaryotic genes are translated with vastly different rates even so they are present on mRNA in equamolar amounts. Thus, the genome of the single stranded RNA bacteriophage MS2, an object of the present studies, contains four cistrons but one of them is translated much more efficiently than the other genes. Moreover, the AUG codon that is mainly used as the translational start, unlike nonsense triplets, is nothing special itself and is present also within the genes as usual translational triplet. Why one AUG is used as the translational start codon very efficiently, another poorly and but many more are never used for that purpose?

In nearly every initiation event all components of translational machinery are identical, except for mRNA and thus mRNA must possess some regulatory code that tells the ribosomes where and with what efficiency to start making proteins. Countless studies have been devoted to revealing this code and various data have shown the significance of the so-called SD-sequence, ribosomal S1 protein binding sites and the strength of the RNA secondary structure that may occlude a translation initiation region as well as other specific TIR properties have been suggested. Still, uncertainty surrounds the features in the mRNA that control the efficiency of translation initiation and as yet it is not possible to predict with a reasonable certainty the quality of a translation initiation region (TIR). More data on properties of various TIRs are required to improve our understanding of the translational control. Also, most of the experimental approaches only distantly resemble the natural conditions and cannot answer the question on the biological significance of the proposed mRNA regulatory features.

Single stranded RNA bacteriophages like MS2 are convenient objects to study translational control because (i) they lack process of transcription and thereby their gene expression is controlled solely at the translational level. (ii) they easily adapt to mutational perturbations if allowed to propagate in bacteria and thus the biological significance of RNA properties of interest can be analyzed through natural evolution.

The specific aims of the performed experiments are stated in the each presented manuscript and also given in the section 'Results and Discussion'. In overall they were as follows:

- To define specific phage RNA properties essential for the efficiency and control of translation initiation of certain genes on RNA or to assess the biological significance of known or suspected regulatory RNA features on live phage.
- To assess using natural evolution, which mRNA properties within a ribosome binding site are of general relevance for the prokaryotic translation initiation.

LITERATURE REVIEW

The mechanism of translation initiation

Translation is a complex process comprising three stages: initiation, in which the translational machinery is assembled on mRNA; elongation, in which amino acids are added to the growing peptide; and termination, in which the nascent peptide is cleaved from the final tRNA and the translational machinery is disassembled. The translational efficiencies in *Escherichia coli* are generally determined at the stage of initiation that requires the 30S and 50S ribosomal subunits, mRNA, fMet-tRNA, three initiation factors, and a GTP molecule. The kinetic model (Fig.1) for initial steps of protein biosynthesis has been elaborated by Gualerzi & others (Gualerzi *et al.*, 1977; Gualerzi & Pon; 1981; Pon & Gualerzi, 1986; Gualerzi & Pon, 1990). During initiation the 30S ribosomal subunit, mRNA, fMet-tRNA, three initiation factors, and a GTP molecule combine to form 30S initiation complex (IC), in which mRNA initiation codon is basepaired with initiator fMet-tRNA anticodon. The 30S IC is the key component in the initiation pathway. The rate of its formation limits the rate of the whole translation process and depends on properties of mRNA TIR.

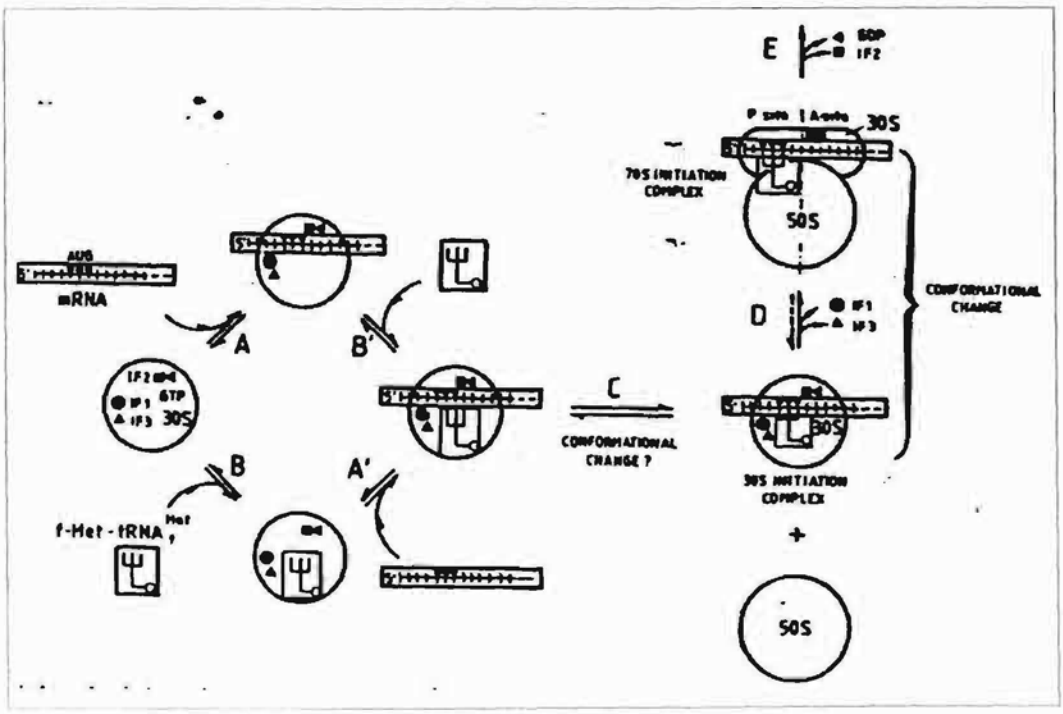


Figure 1. The proposed mechanism of translation initiation in prokaryotes (from Gualerzi & Pon, 1990). According to the proposed mechanism the 30S ribosomal subunit charged with the three translation initiation factors binds in a random order the mRNA and fMet-tRNA-fMet and form a pre-ternary complex in which mRNA and initiator f-Met-tRNA do not interact yet. The codon-anticodon basepairing leads to the formation of 30S initiation complex. The formation of the 30S initiation complex (step C in the figure) is the most important rate-controlling step. The 30S initiation complex can either dissociate or become transformed into the 70S initiation complex by binding of a 50S subunit. The 70S initiation complex is virtually irreversible and enters elongation.

The two crucial steps leading to formation of 30S IC are selection of mRNA by ribosomes and selection of initiator-tRNA and initiation codon (Hartz *et al.*, 1989; Hartz *et al.*, 1991).Ribosomes themselves neither distinguish elongator tRNAs from the initiator tRNA nor recognize the initiation codon (Hartz *et al.*, 1989). The selection of initiator fMet-tRNA over elongators and the verification of the fidelity of the codon-anticodon basepairing are accomplished by initiation factors IF2 and IF3. The codon-anticodon interaction is necessary and essential but not sufficient for the selection of the 'correct' mRNA region where translation should begin. There are a lot more AUG, GUG, UUG codons within the genes (the three most frequent translational starts in *Escerichia coli*) than the actual translational starts.

In contrast, 30S subunit does not need either initiation factors or fMet-tRNA to bind mRNA (van Dieijen et al., 1978; van Duin et al., 1980; Hartz et al., 1991) and find proper initiation regions on it (Hartz et al., 1991, Ringquist et al., 1993). The ribosome, by itself, is able to distinct a good mRNA initiation region from a poor one by the means of sequence specific interactions with mRNA. Presumably, a 30S ribosomal subunit binds to an (unfolded) RNA region in a sequence independent manner and this event trigger the binding of fMet-tRNA or it may be present on ribosome prior the binding. The complex either dissociates from RNA or establishes additional contacts through specific ribosome-RNA and/or codon-anticodon interactions.

mRNA features that establish the efficiency of translation initiation

In *Escherichia coli* the intrinsic efficiencies among different initiation sites can vary 1000-fold (Gold *et al*, 1981; Stormo, 1987). Because in nearly every initiation event all components of translational machinery are identical, except for mRNA, the efficiency with which 30S initiation complex forms depends solely on the properties of the mRNA. These properties include sequence-related features for specific interactions with ribosome and/or initiator tRNA and the secondary structures.

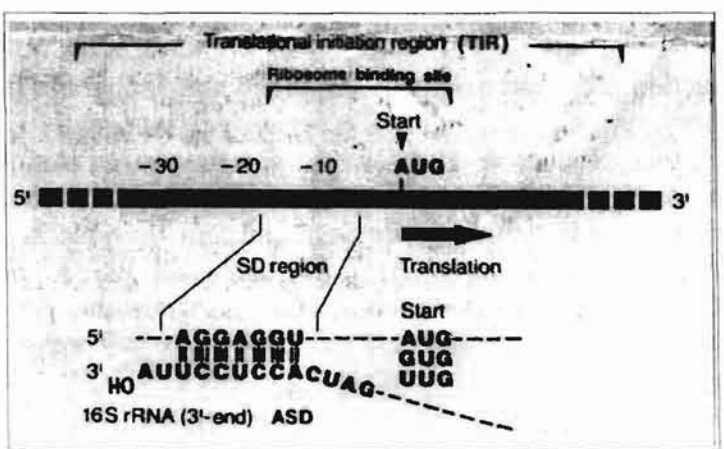


Figure 2. The prokaryotic translation initiation region.
(From McCarthy and Gualerzi, 1990) See text for details

Steitz (1969) was the first person who examined the features of mRNA surrounding true initiation sites. In her experiment 70S ribosomes in the presence of fMet-tRNA were bound to genomic RNA of the phage R17, the RNA was degraded with RNase and the mRNA regions protected from degradation were isolated and sequenced. These ribosome-protected regions are termed ribosome binding sites (RBS) (Fig.2). The regions that lie outside the RBS also can affect translation initiation. To describe the key elements that define efficiency of a translational start the term translation initiation region (TIR) was introduced (McCarthy & Gualerzi, 1990). The TIR always includes the RBS but usually expands beyond it.

The three RBS isolated by Steitz (1969) were of about 30-35 five nucleotides and contained the initiation AUG codon near the middle. No other similarity between the isolated fragments was noticed at that time. Translation initiation codon in majority of cases (90%) is AUG but GUG (8%) and UUG (1%) also function in a number of messages (Gren, 1981; Gold, 1988, McCarthy & Gualerzi, 1990). mRNA encoding IF3 contains AUU as translational start (Sacerdot *et al*, 1982). fMet-tRNA is used with every initiation codon. AUA, AUC (Romero & Garcia, 1991) also allow inefficient translation but to our knowledge are not used as translational starts in the native prokaryotic messages. The translational efficiency of the common start codons decreases in the order AUG>GUG>UUG (Looman & Knippenberg, 1986; Khudyakov et al., 1988; Adhin & van Duin, 1989; Rinquist *et al.*, 1992) and thus certain codons can be a simple mean to affect the initiation rate. The use of rare codons as UUG and AUU in natural messengers seems to serve translational control purposes rather than simple inhibition of translation initiation. For instance, in RNA bacteriophage fr the lysis gene starts with UUG codon but the codon is not utilized by ribosomes in *de novo* initiation. Instead, the translation of the L gene requires the termination of the upstream coat cistron on a nearby located coat gene stop codon. It is the reinitiation at the UUG rather than direct ribosome binding that triggers the lysis translation. When UUG was substituted by GUG or AUG the control was lost because the L gene became expressed independently (Atkins *et al.*, 1989).

Another sequence present in vast majority of prokaryotic TIRs is purine-rich stretch at a variable distance (mostly 7-9 bases) upstream the initiation codon. This so-called Shine-Dalgarno SD-sequence is complementary to 3'-end of the 16S ribosomal RNA, with which it interacts by means of RNA-RNA basepairing during initiation. The interaction proposed originally by Shine and Dalgarno (1974) on the basis of sequence analysis has gained direct experimental support (Hui & deBoer, 1987; Jacob *et al.*, 1987; deBoer & Hui, 1990). The longer interaction seems to permit better initiation efficiency (Ringquist *et al.*, 1992). The presumed principial functions of SD-sequence as compiled by de Smit and van Duin (1994) at least on structured messengers, are as follows. (i) It anchors the 30S subunits in the neighborhood of the translational start site. (ii) It provides the ribosome with sufficient affinity for the available ribosome binding sites. (iii) It increases the ability of ribosomes to compete against the secondary structures (see below) that may be present in the messenger. The spacing between the SD-sequence and the translation initiation codon influence the efficiency but within range 5-13 nucleotides the effects are small (Gold, 1988; Ringquist et.al., 1992).

The secondary structure of mRNA TIR is important determinant of the initiation rate because can hide the SD-sequence from inspection by ribosomes. Stormo (1987) predicted that there must be quantitative relationship between the strength of the secondary structure of the initiation region and the translational efficiency. deSmit & van Duin (1990a,b) shown it to be true. They varied the strength of a hairpin that contains the RBS of the phage MS2 coat protein gene. The results revealed strict correlation between translational efficiency and stability of the hairpin. Increase of the stability of the structure by -1.4 kcal/mol give rise the reduction of translation initiation rate by a factor of 10. On the other hand the stability of wild type hairpin appeared to allow unabated translation.

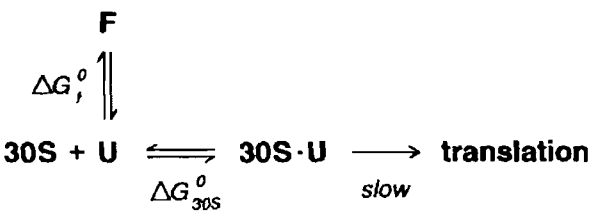


Figure 3. Schematic representation of the competition between hairpin formation U=unfolded, F=folded) and binding to 30S ribosomal subunits. ΔG_f is the free energy of helix formation, ΔG_{30S} is the free energy of binding of a 30Ssubunit to the unfolded RNA. slow indicates the rate-limiting step in translation (from de Smit and van Duin, 1990a).

The mathematical analysis of the obtained data led the authors to conclude that the efficiency of translation initiation depends on fraction of molecules in which the ribosome binding site is unfolded and that ribosomes bind only unfolded initiation sites. The weaker is the structure the more time the RBS is available for ribosome binding. On the other hand the ribosome binding to the unfolded mRNA is dependent on the affinity of the initiation region for ribosomes, for instance, the strength of SD-interaction. The higher is the affinity of ribosomes for the initiation site the higher is probability that the unfolded molecule will be bound by ribosomes. The translation initiation on structured messengers can therefore be described as a result of continual competition between formation of the helix and binding to a ribosome (Fig.3).

Interestingly, in *in vitro* studies it has been found that a SD-interaction is mechanistically irrelevant for correct initiation (Calogero et al., 1988). The authors proposed that under natural conditions the SD-sequence play a crucial role in the competition between messengers for limited pool of free ribosomes. A support for such view comes from the finding that the genes bearing mutations inside the SD-sequence or initiation codon experience more significant decrease in their expression as the genes mutated outside these regions under conditions of growth in minimal medium (Jacques et al., 1992).

Another ribosome-mRNA interaction that plays role in the translation initiation is interaction of ribosomal protein S1 with mRNA. It unfolds mRNA secondary structures and is particularly required for translation of highly structured mRNAs (van Dieijen *et al*, 1976; Subramanian, 1983). S1 protein has a preference for pyrimidine-rich sequence. It has been shown to bind to U-rich sequences present

upstream from the start of the coat protein gene of the single stranded RNA bacteriphages fr and Qβ. Such U-rich sequences are present in the vicinity of many efficient E.coli translational starts (Boni et al., 1990). Runs of pyrimidines may act as enhancers for translation initiation.

Within the prokaryotic genes there are many SD-like sequences followed at a correct distance by AUG or GUG. This shows that some additional source of information pinpoints the genuine starts. Statistical analysis (Sherer *et al.*, 1987) showed that distributions of nucleotides within the ribosome binding is non-random at several positions other than initiation codon and SD-sequence and in overall the ribosome binding sites are rich in A and also U. One possibility is that the information pattern reflects an attempt to keep ribosome binding site free of secondary structure. On the other hand ribosomes might have a preference for certain nucleotides or sequences. Other sequences within RBS have been proposed to interact with 16S rRNA but they either lack an evidence in recognition of a TIR through rRNA-mRNA interactions or the experimentally data are contradictory (reviewed by McCarthy and Gualerzi, 1990; Sprengart and Porter, 1997). At least some of these so-called translational enhancers may serve as binding sites for S1 protein (Tzareva *et al.*, 1994; Ringquist *et al.*, 1995).

The single stranded RNA bacteriophages

The single stranded RNA bacteriophages (Zinder, 1975) such as MS2 contain single-stranded RNA genome encapsidated in a shell, composed of 180 copies of coat protein and a single molecule of the A protein. Basing on several criteria they are classified in four groups. Groups I and II are similar and collectively referred as group A and also groups III and IV are closely related (group B). Phage MS2 is representative of the first group. Its single stranded RNA genome (see at Fig.5) is 3569 nucleotide long and contains four genes. The maturation (A) protein is present as one copy per virion and is required infection. The coat (C) protein is the major structural component of virion, the product of the lysis (L) gene is needed for cell lysis and the replicase (R) protein serves RNA replication. Group B phages lack separate lysis protein (lysis function these phages performs maturation protein) and contain additional readthrough protein.

Infection cycle

The infection cycle of an RNA phage have been studied on group B phages and is depicted in Figure 4 (Biebricher & Eigen, 1987; as adapted by Olsthoorn, 1996). Phages attach to the F pili of the E.coli cell *via* maturation protein and eject its RNA into the bacteria through the central channel of the pilus. Once the (plus)-strand RNA genome is inside the cell, the phage development starts. The level and timing of phage gene expression is carefully controlled and tightly coordinated with replication.

After entry the cell RNA first serves as messenger for the coat gene translation which RBS is the only one available to ribosomes on intact phage RNA. Translation initiation regions of other three genes are hidden by phage RNA secondary structure. Ribosomes translating the coat message disrupt a secondary

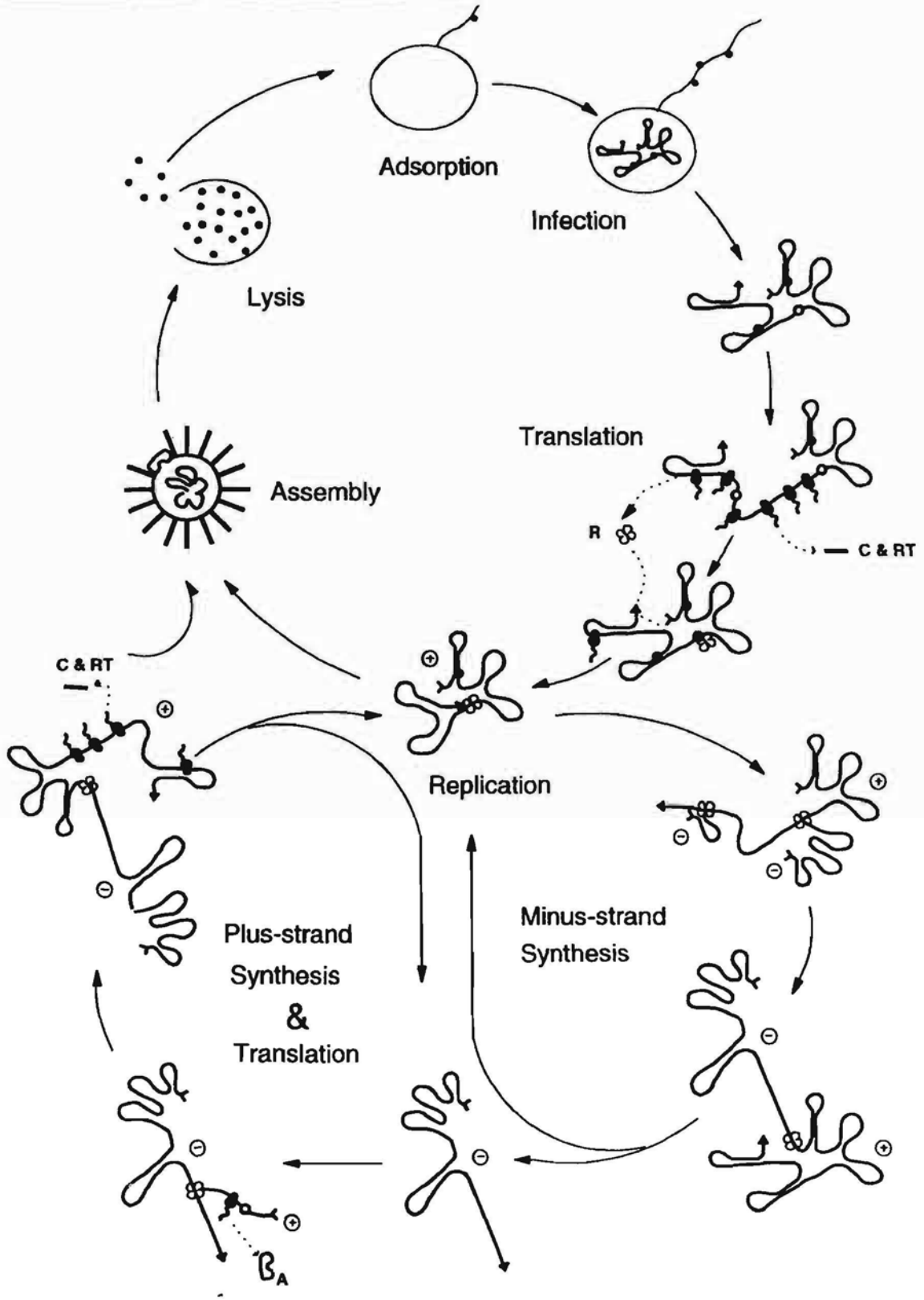


Figure 4. Infection cycle of an RNA phage. R-replicase subunit, C-coat protein, A-maturation protein, O-open ribosome binding site (RBS), ● - closed RBS.

structure that occludes the initiation site of the replicase cistron. The replicase protein is produced and replication may start.

Phage RNA replication involves a protein complex, the replicase holoenzyme, that consists of the phage-encoded replicase subunit and for proteins of the host cell, elongation factors EF-Tu and EF-Ts, ribosomal protein S1 and a protein called host factor (HF). It has been shown that replication cannot occur on an RNA molecule that is being translated (Kolakofsky and Weismann, 1971). Before the replication begins the phage RNA must be cleared from ribosomes. The replicase holoenzyme attaches to the initiation region of the coat protein gene. The binding most probably mediated by S1 protein (Boni *et al.*, 1990) that is a component of both ribosome and replicase holoenzyme. Since the coat RBS is the only one accessible for direct translation initiation the replicase binding to the coat TIR prevents the ribosome binding to the phage RNA and replication can proceed.

Replication proceeds via synthesis of a free plus and minus strands. Since no double-stranded RNA replication intermediate occurs, proteins can be synthesized from nascent RNA plus-strands. During synthesis of plus strands, the RBS of the maturation protein is temporarily accessible and small amount of maturation protein is produced. The lysis protein is synthesized at a low level throughout the infection and its translation is coupled to that of the coat gene. As the coat protein concentration rises it binds to the replicase start region and physically excludes ribosomes from further initiation of the replicase synthesis. At about 20 min post infection the replicase synthesis is fully shut off. As the concentration of phage structural components builds up phage assembly begins. The phages are released from bacteria after disruption of the cell wall by the phage lysis activity.

Control of the gene expression

The coat protein gene is abundantly expressed throughout the infection to satisfy phage needs for huge amounts of the protein. The three other phage-encoded proteins are required in small quantities and their synthesis is temporarily regulated. A major role in control is played by phage RNA secondary structure which can deny access of ribosomes to translational start regions (reviewed by van Duin, 1988; de Smit & van Duin, 1991).

The maturation protein gene

The maturation (A) protein is needed only one copy per virion and its synthesis accordingly is kept at a low level. The translation of the A protein gene is restricted to nascent plus strands. In completed RNA chains its translational start is hidden by secondary structure that arises from a basepairing between the RBS of the A protein gene and an internal A protein gene coding sequence. When however the phage RNA is being replicated, the inhibitory folding needs some time to be formed. This delay in folding allows the ribosomes afford the binding to the TIR of the maturation gene and thus to produce the required amounts before the start is shout down by the secondary structure. Such a model has been in literature for a long time (Fiers *et al*, 1975; Robertson & Lodish, 1970) and recently was shown to be correct (Groeneveld *et al.*, 1995; Poot *et al.*, 1997).

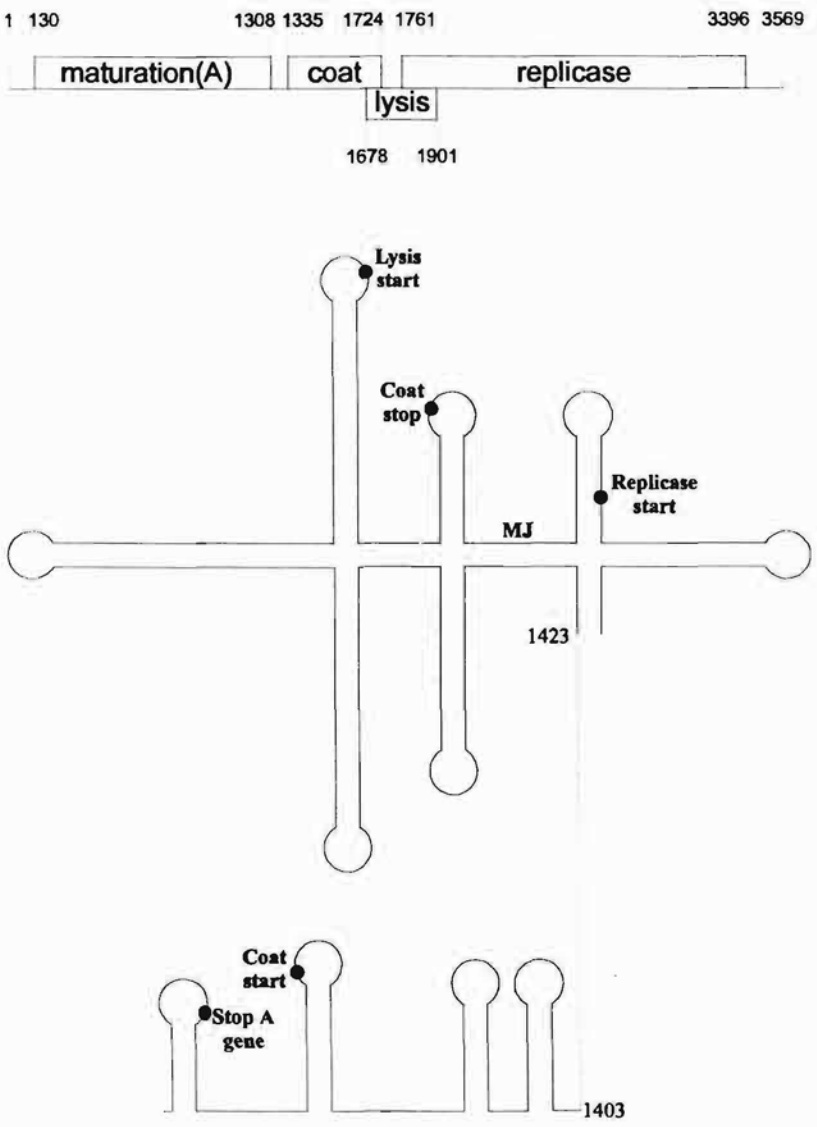


Figure 5. Genetic map of RNA phage MS2 and schematic drawing of the MS2 RNA secondary structure of the central part of MS2 RNA (adapted from Groeneveld, 1997). The translational start and stop codons are indicated. The dotted line marks RNA region which structure is uncertain.

The coat protein gene

Except for the competition between ribosome and replicase for the ribosome binding site described above the synthesis of the coat protein is not negatively regulated in any way. The coat gene RBS (de Smit & van Duin, 1991) that adopts a simple hairpin structure (see p.17 in RESULTS) with the start codon in the loop and the SD-region in the stem. This secondary structure in native sequence context does not interfere with the efficiency of translation initiation. Destabilization of the stem beyond wild type level did not lead to increased protein yields whereas stabilization of hairpin, no matter how achieved, causes translation to fall.

It appears that the phage RNA is designed in such a way to maximally activate the coat gene RBS. The phage genome region of some 1200nt that precedes the coat gene RBS about 30 times stimulates the coat protein synthesis as compared to the situation when it is absent or substituted by the phage-unrelated sequence. The progressive shortening of this region leads to gradual decrease in coat protein synthesis. Thus, in these truncated constructs the stability of the hairpin restricts the ribosome access to the coat gene start. It can be said that translation initiation region of MS2 coat gene expands over thousand or so nucleotides. How the upstream RNA achieves this stimulation is not evident but de Smit & van Duin (1994) provided several possible explanations.

The lysis protein gene

The translation of the coat gene is needed to produce the lysis protein (Schmidt *et al*, 1987). This translational coupling is enforced by a strong hairpin structure that involves the RBS of the lysis gene and prevents ribosome binding to the translational start. The current model is that some ribosomes terminating the coat translation reach the L start by random lateral movement along the mRNA and re-initiate at the lysis start. Interestingly, the aminoacids encoded by coat-lysis overlap and the coat-replicase intercistronic region are dispensable for the protein function that hints on a merely regulatory role of the 'unnecessary' gene extension. The biological significance of the coat/lysis coupling that probably lies within both preventing the cell disruption unless the virion formation has reached a sufficient height and the above described switching between phage RNA translation and replication has been demonstrated by Klovins and others (1997).

The replicase protein gene

The control of the replicase gene expression is rather sophisticated. First, translation of the upstream coat gene is needed to activate the ribosome binding site of the replicase gene. As the concentration of coat protein rises it binds to the specific hairpin structure that adopts the initiation site of the replicase gene and represses translation again. Both the molecular mechanism of the coat-replicase translational coupling and especially the interaction of the phage coat protein with the replicase.

The translational coupling between the coat protein and the replicase gene was deduced from the observation that early amber mutation at codon 6 of the coat gene prevents the replicase synthesis whereas late amber mutation at codon 50 does not have such an detrimental effect on the replicase gene translation (Horiuchi, 1975). Once MS2 RNA sequences became known the first structural model to rationalize these observations was proposed (Min Jou et al., 1972). In this model the coat gene 1409-1423 coat gene sequence was postulated to basepair with the replicase ribosome binding site. Berkhout & van Duin (1985) further analyzed the control mechanism using MS2 cDNA clones and confirmed the MJ-model. It was found that deletion of the proposed coat gene regulatory segment uncouples replicase translation. This and further analysis of the MS2 RNA secondary structure (Skripkin et al., 1990) led to suggest the long-distance interaction of 6-7 basepairs as the molecular basis for the control mechanism (see p.18 in RESULTS). This MJ-interaction was later proved to be correct exploiting basepair substitution analysis on partial MS2 cDNA clones (van

Himbergen et al., 1993). The MS2 RNA structure model presented in Results is an improved version (Groeneveld, 1997) of the Skripkins' model and in the region of the replicase TIR differs by the presence of VD-interaction that in the oldest model was absent.

The other control mechanism operates later in infection. The coat protein of the bacteriophage interacts with the stem-loop structure, often designated as translational operator, that adopts the replicase translation initiation region and represses the translation of the replicase cistron (Bernardi & Spahr, 1972). This interaction has been extensively studied and the structural requirements in RNA for efficient binding are known in details (Uhlenbeck *et al.*, 1983; Romaniuk *et al.*, 1987) (see p.18, RESULTS). The exact secondary structure is essential but the sequence is not. The hairpin must contain the bulged purine at specific position and the ANYA tetraloop (N-any nucleotide, Y-pyrimidine). The location of this translational operator is conserved and its affinity for coat protein is conserved among RNA phages of various groups, although the structural requirements are specific for representatives of different groups (Witherel *et al*, 1991). This same interaction was thought to be necessary for phage RNA encapsidation (Beckett & Peabody, 1988) but Peabody demonstrated recently (1997) that the coat protein interaction is not required for encapsidation of the genome.

Assay of the formation of translation initiation complex *in vitro*

The filter-binding assay was used to measure the activity of template oligoribonucleotides – analogs of the replicase gene TIR in 70S ribosome binding. RNA was incubated with dissociated ribosomes in the presence of purified translation initiation factors and ^{35}S-labeled fMet-tRNA. The mixture was then filtered through nitrocellulose filter. The ribosomes and the initiation complexes are retarded on the filter. The amount of ribosome-bound ^{35}S label is proportional to the amount of ribosome-bound template RNA (Berzins *et al.*, 1988).

Evolution experiments

The evolutionary approach (Olsthoorn, 1996, Olsthoorn *et al.*, 1994) was used to reveal both the mRNA features essential for translation initiation and the biological significance of certain regulatory phage RNA structures. It rests on the lack of a proofreading activity of the phage RNA polymerase (the replicase). As a consequence the misincorporation frequencies in RNA phage genomes are very high (10^{-3}– 10^{-4} per nucleotide per replication (Batschelet *et al.*, 1976)) and these result in an extremely heterogeneous RNA phage population (Domingo *et al.*, 1978). Natural selection excludes the non-viable variants and leads to the survival of good ones. Because there is a constant 'supply' of deviant virus genotypes by the error-prone replication new variants are continuously tested by natural selection and as a result RNA viruses may change very rapidly and adapt to new conditions and mutational perturbations with amazing easy.

The experimental approach consists in modifying a known or suspected RNA sequence/structure feature by mutational perturbations that reduce phage viability without abolishing it completely (at least in a sense that the phage can yield viable progeny by the error-prone replication). Progeny phages are then observed through number of growth cycles to evolve back to high viability and analyzed with regard to the accumulated spontaneous mutations compensating for the initial perturbation. Those containing such compensatory are often called pseudo-revertants.

Directed mutations, either distinct or random, were introduced in phage MS2 RNA genome *via* its full-length cDNA copy that is present on a plasmid behind the inducible P_L promoter of the phage. The advantage of 'random' approach is that it allows generate many different cDNA variants of the RNA feature under scrutiny in a single step. The plasmid yields viable phage and is maintained in *E.coli* F^- host cells. The phages are released from bacteria but the host cells cannot be re-infected because they lack F pili. The produced mutants are then used to infect *E.coli* F^+ bacteria and their evolutionary adaptations by suppressor mutations are recorded. Typically, many mutants of an RNA property are monitored that allows to reveal common rules and essential features in an evolutionary analysis.

***In vivo* measurements of the replicase gene expression**

To assess the effect of both initial and compensatory mutations on the replicase gene translation cDNA fragments of the phage genome were cloned in an expression vector (van Himbergen *et al.*, 1993) where the *lacZ* gene is fused to the 3'-portion of the replicase gene (RlacZ). The replicase expression was measured as the β-galatosidase activity of RlacZ fusion protein by a conventional method (Miller, 1972). In that method the β-galatosidase activity is revealed from the extent of breakdown of its colorless substrate ONPG that is converted by the enzyme into a yellow-colored product. The intensity of the color depends on the concentration of the enzyme in the sample.

RESULTS AND DISCUSSION

In vitro data

The ribosome binding site on bacteriophage MS2 RNA contains translational start codon AUG near the middle and putative SD-sequence at a distance of 6-7 nucleotides. There are two variants for the SD-interaction between the replicase gene TIR and 16S rRNA. One involves the mRNA template sequence 5'-UGAGGA-3' that can basepair with the very 3'-end of the 16S rRNA - 3'-AUUCCU-5', the other is interaction of the 5'-GAGG-3' sequence with the 3'-CUCC-5' region of rRNA. It is not clear which one takes place during translation initiation in the cell.

1. AAACAUGAGGAUUACCC**AUG**UCGAAGACAACAAAG
2. AAACAUGAGGAUUACCC**AUG**
3. CAACAUGAGGAAUACCC**AUG**
4. AACAUGAGGAAUACCC**AUG**
5. AUGAGGAAUACCC**AUG**
6. UGAGGAAUACCC**AUG**

Figure 1. The RBS of the phage MS2 replicase gene (1), the RNA fragments of MS2 and fr that retain full template activity (2-5) initiation site for phage MS2 and the minimal sequence exerting a reasonable ribosome binding (6). The replicase gene start codon is in bold, a variant of SD-sequence is underlined.

Previous studies on MS2 and related phage fr (Berzin *et al.*, 1978; Borisova et al., *1979*; Berzin *et al.*, 1982; Renhof *et al.*, 1985) had led to show that the fragments in length of 20 nucleotides (Fig.1, sequences 2 and 3, respectively) exert the same functional activity in ribosome binding as the whole RBS. Thus all of the information essential for efficient initiation of replicase synthesis is contained within 17 nt region preceding the translational start. MS2 and fr RNA differs at two positions in the shown sequence but these deviations have no influence on affinity of mRNA for ribosome.

In order to determine more precisely the 5'-boundary of an active replicase gene TIR progressively shortened templates (Fig.1, templates 4-6) were used in filter binding assay. The results showed that 16 nt long fragment (template 5) retains full activity in driving the formation of the 70S-initiation complex. The shortest of used variants exerted lowered but still reasonable template activity (sequence 6).

The mutational analysis revealed the UGAGGA sequence as that responsible for ribosome binding with high affinity. The substitutions in this region leaves the template no or only a low activity. More precise mapping would be needed to dissect the effects of particular nucleotides or short sequences contained within this fragment on the efficiency of ribosome binding.

The essential conclusion from *in vitro* data applied later in evolutionary studies is that mutations outside UGAGGA sequence and AUG codon should not directly in a sequence specific manner affect the efficiency of the initiation of the replicase protein synthesis. On the other hand the elimination of the UGAGGA sequence must prevent efficient translation of the replicase gene.

Translation initiation region of the phage coat protein gene

The start of the coat protein gene adopts a simple hairpin structure (Fig.2). The data of expression studies on partial MS2 cDNA clones had indicated (de Smit & van Duin, 1990) that the stability of this hairpin is important for the coat gene expression. Any mutation that increased the strength of the coat initiator hairpin reduced the coat protein yield. On the other hand the destabilization of the structure had no effect on coat gene translation. Thus, the strength of the hairpin appeared just weak enough to unabated translation and it suggested that the hairpin has evolved the highest stability compatible with maximal ribosome loading.

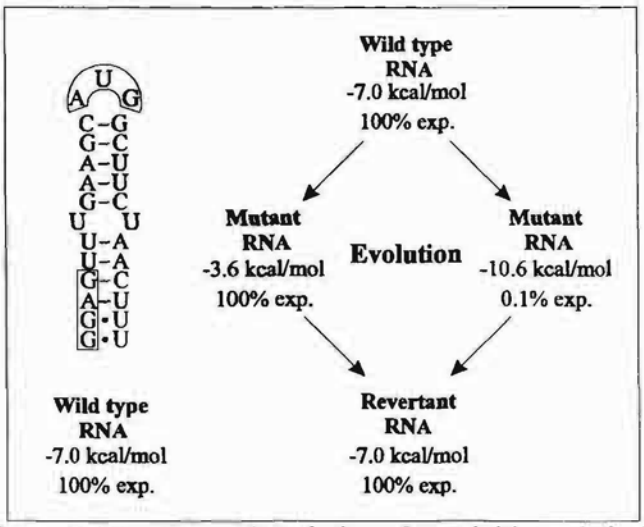


Figure 2. Structure and evolution of coat initiator hairpin.

To examine the presumed tuning of this helix stabilizing and destabilizing mutations were introduced in live phage and the evolution of mutants to high viability was monitored. It was observed that both types of mutations are strongly selected against and the phage fast reverts to hairpin stability that is close or the same as that of wild type without adopting the native sequence (Fig.2). These results provide evidence on the biological significance of the presumed tuning of the stability of the coat initiation hairpin. The return of the stabilized hairpins to the wild type stability arises from the need to synthesize adequate amounts of coat protein. The reason why destabilized hairpins are also selected against is not presently understood. The results also show that except for the translational start codon and the SD-sequence the efficiency of the coat gene translation initiation is not affected by primary sequence of the RBS but strongly depends on the strength of its secondary structure.

Translation initiation region of the replicase protein gene

The replicase protein is limited to only a small burst early in infection. The negative control is accomplished by repressor – phage coat protein binding to the replicase initiator hairpin and long-distance MJ-interaction between a coat gene segment and replicase translational start region (Fig.3).

(i) By site directed mutagenesis we randomized the 1746-1760 sequence of the replicase TIR in phage RNA. With respect to the initiation and control of the replicase gene translation such mutations have dual effect. First, because the mutated region includes the SD-sequence, the base changes should turn the initiation region into very inefficient one. Thus, to ensure an adequate supply of replicase for appropriate RNA replication the phage RNA mutants must somehow activate the replicase TIR upon evolution. It was expected to reveal the RNA features of general relevance for translation initiation. Second, the introduced substitutions destroy the native secondary structure of the initiation region and the binding of the coat protein to the replicase TIR with high efficiency is impossible for such phage RNA variants. It was not known whether a native-like initiator hairpin of the replicase gene is required for the encapsidation of the phage RNA. And if the phage can survive without it how the pseudorevertants will adapt the efficiency of the replicase gene translation to the absence of repression.

(ii) Similar 'random' approach was applied to assess the biological significance of the MJ-interaction for the control of the replicase gene translation. We reasoned if the LDI is essential for live phage the mutants, in which RNA the MJ-interaction is broken and consequently the replicase synthesis is activated, should evolve suppression of the replicase translation again.

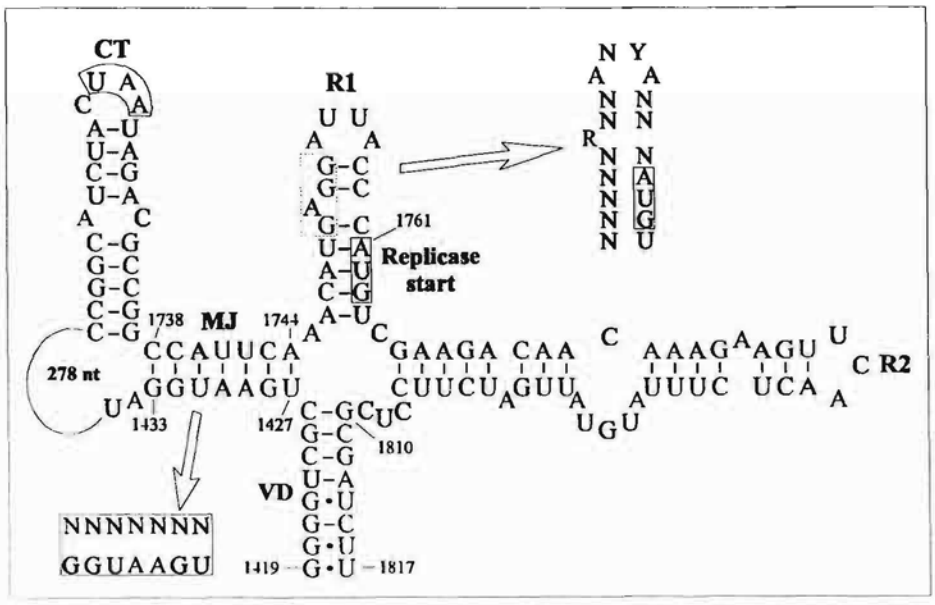


Figure 3. The secondary structure of the replicase gene TIR. The translational start AUG codon, a variant of SD-sequence as well as the coat gene stop codon are marked by box. The regions that were randomized are indicated. R1, R2 and CT designates local stem-loop structures. MJ designates the MJ –interaction, VD designates the 1419-1426/1810-1817 LDI.

Phage RNA mutants of SD-sequence and replicase gene initiator hairpin

The results on properties and evolution of mutants of SD-sequence and initiator hairpin of the replicase gene are presented in Figure 4.

- The many obtained MS2 mutants did not share RNA sequence (SD-box) and structure (an RNA operator hairpin) features characteristic for the native TIR of the replicase gene. They also lacked some other SD-sequence in the randomized region and the replicase synthesis dropped drastically.

The mutations decreased the phage titer by two and more orders of magnitude but did not prevent the phage production obviating that the replicase synthesis is not abolished in full. This is consistent with observations that SD-sequence is not an absolute requirement for translation initiation (Calogero et al., 1988; Melacon *et al.*, 1990; Boni et al., 1991). Another indication given by properties of mutants is that an AUG codon preceded by non-conventional for the initiation sequences may yield some leakage of translation depending on the context of the surrounding sequence (presence or absence of SD-like sequences AAG, GG etc) and, especially, structure.

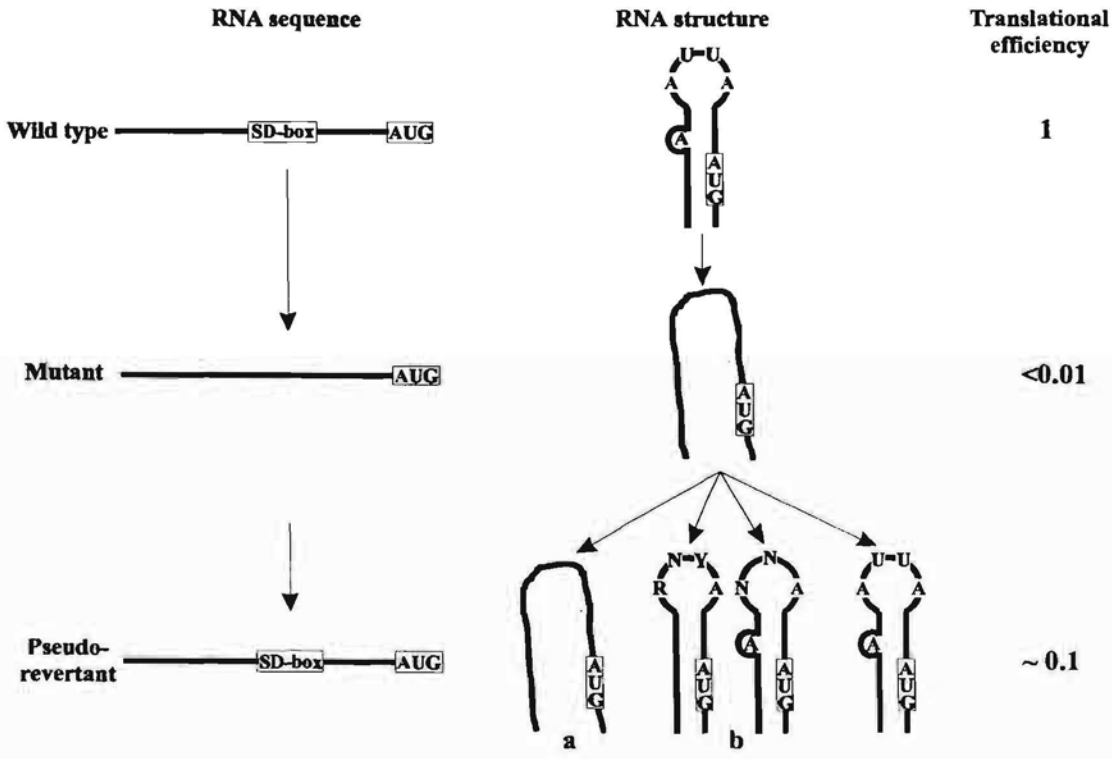


Figure 4. Properties and evolution of mutants of the replicase geneTIR.

- The evolution of phages followed a straightforward sequence pattern. Thus, out of 28 suppressor mutations (Firg.5) evolved by various pseudorevertants more than half resulted in guanosine nucleotides. The sequence analysis revealed that this biased content of the suppressor mutations originates in creation of various SD-regions in the replicase initiation site.

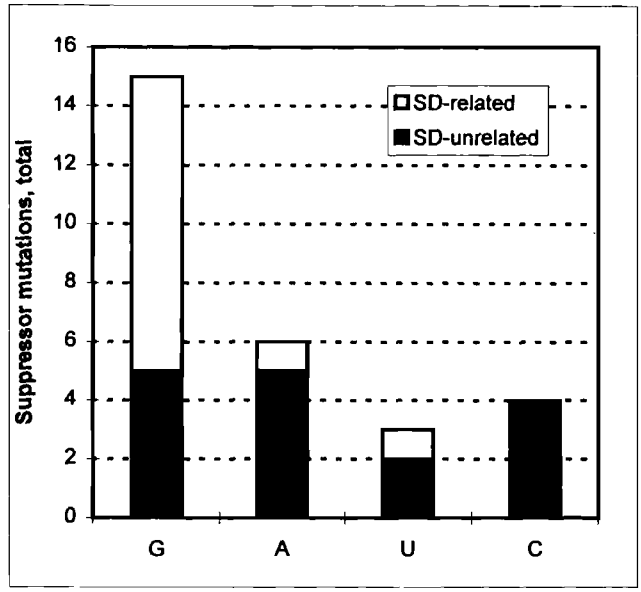


Figure 5. The biased content of the suppressor mutations.

In the above evolutionary analysis the SD-region clearly emerged as the main positive determinant of a prokaryotic TIR. This finding is in line with well-known role of SD-sequence in translation initiation (Gold, 1988; Dreyfus, 1988; de Smit and van Duin, 1944). We found neither any other common sequence-specific pattern nor a pronounced nucleotide bias, except for SD-related guanosine nucleotides. This shows that the specificity of suppressor mutations *per se* has played no significant role in evolution of replicase TIR from false to true.

- Another distinct property of most of the adapted phage RNA variants was the lack of a translational operator hairpin (secondary structure models a and b, Fig.5), whereas a feasible structural feature in number of instances was evolution of partially resembling structures (hairpins b, Fig.4). We suggest that such hairpins may serve as a low affinity binding sites for phage coat protein.
- In either of the examined cases the replicase gene translation in pseudorevertants was relatively high as compared to their progenitors but relatively low as compared to wild type. The yield of replicase did not correlate with the 'strength' of SD-sequence and the access of the replicase gene start was strongly influenced by phage RNA structure. It is conceivable that the limited activation of the replicase production represent a phage adaptation to the absence of efficient repression of the replicase gene translation by phage coat protein.

Translational coupling

Manuscript 5 describes a study intended to reveal whether the MJ-interaction possesses the biological significance for the control of the replicase gene translation. The answer turned out to be positive. We found that phage mutants, in which RNA the MJ-interaction was destroyed, evolved the suppressor mutations placing the replicase gene under translational control again. For some of the pseudorevertants lacking the MJ-interaction the exact molecular basis for the renewed control is not evident. Yet, one particular suppressor mutation U1817C that was acquired by the mutant phage RNA most frequently indicated the regulatory implication of another RNA-RNA long-distance interaction, the 1419-1426/1810-1817 (VD-interaction) LDI, in translational control of the phage MS2 replicase gene.

A principal support for this possibility was obtained later and is summarized in Figure 5. The mutations that disrupt the feasible control LDI but preserve the encoded aminoacids were introduced in the coat gene and cloned both in an expression vector and infectious MS2 cDNA copy. In the expression vector the two mutations about twice increase the replicase synthesis as compared to wild type. Another evidence in the role of VD-interaction in the control of the replicase gene translation was provided by two types of pseudorevertants, which arose during evolution of the mutant phage. Several of its progeny phages contained the U1817C suppressor mutation that was dominant in the MJ-analysis (Rev.2) but in one particular case we monitored the exchange of the replicase gene initiation AUG codon to GUG codon (Rev.3) that provide another evidence in the participation of the VD-interaction in the replicase gene control. In agreement with data of others (Looman & Knippenberg, 1986; Khudyakov *et al.*, 1988) we found that GUG reduce the translation as compared to AUG some 2-3 times. Several descendants of the mutant partially reverted to wild type RNA sequence (Rev.1) and they are therefore uninformative.

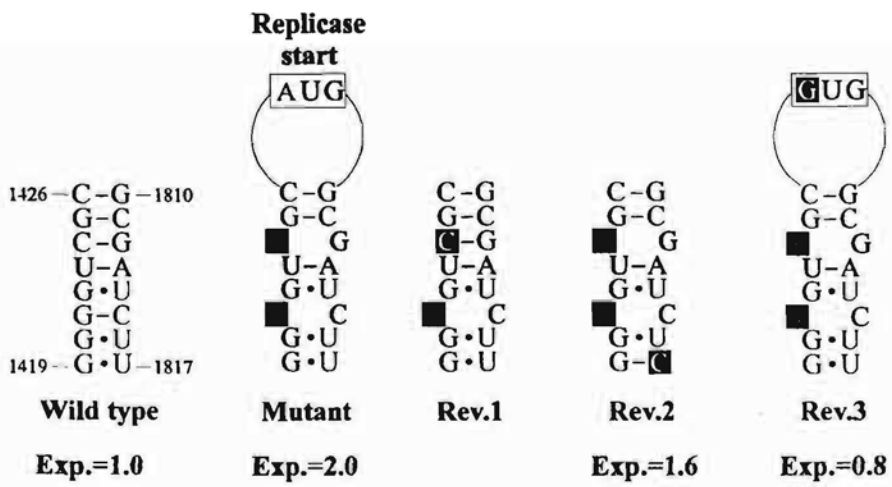


Figure 6. Properties of mutants and pseudorevetants of the VD-interaction. The initial mutations are in gray while suppressor mutations in black boxes. Exp. – the efficiency of the replicase gene expression relatively to the wild type.

Conclusions

1. The essential information for the efficient initiation of the replicase gene translation in phage MS2 and fr RNA is located within a short 16 nucleotide RNA fragment that contains the initiation AUG codon and the SD-sequence.

2. The precise stability of the hairpin that contains the translation initiation codon and the SD-sequence of the phage MS2 coat protein gene is essential for the phage viability.

3. Except for the SD-sequence and translation initiation codon the efficiency of the coat protein gene translation is not affected by other primary sequences within its ribosome binding site.

4. The SD-sequence is the major positive determinant of the translation initiation process in *Escherichis coli* as revealed by evolution of the mutants of the phage MS2 RNA replicase gene TIR. Other primary sequences within the ribosome binding site are not of general relevance for the initiation process, except for the translational start codon.

5. SD-sequence is not an absolute requirement for translation initiation *in vivo*.

6. The global secondary structure of the phage MS2 RNA exerts strong negative effect on the efficiency of the replicase gene translation.

7. The interaction of the phage MS2 coat protein with the ribosome binding site of the replicase gene is not required for the encapsidation of the phage genome.

8. The long-distance RNA-RNA MJ-interaction between the internal segment of the phage MS2 coat protein gene and the start region of the replicase gene is essential for the control of the replicase gene translation. Another LDI between the coat protein and the replicase protein genes participates in this regulation.

REFERENCES

Adhin, M.R. & van Duin, J. (1989). Translational regulation of the lysis gene in RNA bacteriophage fr requires a UUG initiation codon. Mol.Gen.Genet. **218**, 137-142.

Batschelet, E., Domingo, E. & Weissmann, C. (1976). The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene* **1**, 27-32.

Berkhout, B. & van Duin, J. (1985). Mechanism of translational coupling between coat protein and replicaase genes of RNA bacteriophage MS2. *Nucl. Acids Res.* **13**, 6955-6967.

Berkhout, B., Schmidt, B.F., van Strien,A., van Boom, J., van Westrenen, J. & van Duin, J. (1987). Lysis gene of bacteriophage MS2 is activated by translation termination at the overlapping coat gene. *J. Mol. Biol.* **195**, 517-524.

Bernardi, A. & Spahr, P.F. (1972). Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17. *Proc. Natl. Acad. Sci. USA* **69**, 3033-3037.

Berzins, V., Borisova, G.P., Cielens, I., Gribanov, V.A., Jansone, I., Rosenthal, G. & Gren, E. (1978). The regulatory region of MS2 phage replicase cistron. Functional activities of individual MS2 RNA. *J. Mol. Biol.* **119**, 101-131.

Berzins, V., Cielens, I., Jansone, I. & Gren, E. (1982). The regulatory region of phage fr replicase cistron. III. Initiation activity of specific fr RNA fragments. *Nucl. Acids Res.* **23**, 7763-7775.

Borisova, G., Volkova, T.M., Berzins, V., Rosenthal, G. & Gren, E. (1979). The regulatory region of MS2 phage RNA replicase cistron. IV. Functional activity of specific MS2 RNA fragments in formation of the 70S initiation complex of protein biosynthesis. *Nucl. Acids Res.* **6**, 1761-1774.

Calogero, R.A., Pon, C.L., Canonaco, M.A. & Gualerzi, C.O. (1988). Selection of the mRNA translation initiation region by *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. USA* **85**, 6427-6431.

de Boer, H.A. & Hui, A.S. (1990). Sequences within ribosome binding site affecting messenger RNA translatability and method to direct ribosomes to single mRNA species. *Meth. Enzymol.* **185**, 103-114.

de Smit, M.H. & van Duin, J. (1990a). Control of prokaryotic translation initiation by mRNA secondary structure. *Progr. Nucl. Acid. Res. Mol. Biol.* **38**, 1-35.

de Smit, M.H. & 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.

de Smit, M.H. & van Duin, J. (1994). Translation initiation on structured messengers: Another role for the Shine-Dalgarno interaction. *J. Mol. Biol.* **235**, 173-184.

Dreyfus, M. (1988). What constitutes the signal for the initiation of protein synthesis on *Escherichia coli* mRNAs? *J. Mol. Biol.* **204**, 79-94.

Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell*, **13**, 735-744.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, C., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymakers, A., Vandeberghe, A., Volckaert, C., Ysebaert,

M. vande Kerckhove, J., Nolf, F., van de Kerckhove, J., Nolf, F. & van Montagu, M. (1975) A-protein gene of bacteriophage MS2. *Nature*, **256**, 273-278.

Gold, L. (1988). Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Ann. Rev. Biochem.* **57**, 199-233.

Gold, L. & Stormo, G. (1987). Translation initiation. In *Escherichia coli* and *Salmonella typhirium* (Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Shaechter, M. & Umbarger, H.E. eds), vol.2, pp.1302-1307.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S, Swebilius—Singer, B. & Stormo, G. (1981) Translation initiation in prokaryotes. *Annu Rev Microbiol* **35**, 365-403.

Gren (1984). Recognition of messenger RNA during translational initiation in *Escherichia coli*. *Biochimie*, **66**, 1-29.

Groeneveld, H., Thimon, K. & van Duin, J. (1995). Translational control of maturation-protein synthesis in phage MS2. A role for the cinetics of RNA folding? *RNA*, **1**, 79-88.

Groeneveld, H. Secondary structure of bacteriophage MS2 RNA. PhD thesis, Leiden University, Leiden, The Netherlands.

Gualerzi, C.O., Risuleo, G. & Pon, C.L. (1977). Initial rate kinetic analysis of the mechanism of initiation complex formation and the role of initiation factor IF-3. *Biochemistry*, **16**, 1684-1689.

Gualerzi, C.O. & Pon, C.L. (1990). Initiation of mRNA translation in prokaryotes. *Biochemistry*, **29**, 5881-5889.

Gualerzi, C. & Pon, C.L (1981) Protein biosynthesis in procaryotic cell: Mechanism of 30S initiation complex formation in *Escherichia coli*. In *Structural aspects of recognition and assembly in biological macromolecules* (ed Balaban, L., Sussman, J.L, Traub, W. and Yonath, A.), pp.805-826. Balaban ISS Rehovot and Philadelphia.

Hartz, D., McPheeters, D.S., Green, L. & Gold, L. (1991). Detection of *Escherichia coli* ribosome binding at translation initiation sites in the absence of tRNA. *J. Mol. Biol.* **218**, 99-105.

Hui, A. & de Boer, H.A. (1987). Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **84**, 4762-4766.

Horiuchi, K. (1975) Genetic studies on RNA phages. in Zinder, N.D. (ed.) RNA phages, Cold Spring Harbor Laboratory, pp.29-50.

Jacob, W.F., Santer, M. & Dahlberg, A.E. (1987). A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. *Proc. Natl. Acad. Sci. USA* **84**, 4757-4761.

Jacques, N., Guillerez, J. & Dreyfus, M. (1992). Culture conditions differentially affect the translation of individual mRNAs. *J. Mol. Biol.* **226**, 597-608.

Khudyakov, Yu.E., Neplyueva, T.I., Kalinina, T.I & Smirnov, V.D (1988). Effect of structure of the initiatior codon on translation in *E.coli*. *FEBS Lett.* **232**, 369-371.

Klovins, J., Tzareva, N.A., deSmit, M.H., Berzins, V. and van Duin, j. (1997) Rapid evolution of translational control mechanisms in RNA genomes. *J.Mol. Biol.* **265**, 372-384.

Looman, A.C & van Knippenberg, P.H. (1986) Effects of GUG and AUG initiaton codons on the expression of lacZ in *Escherichia coli*. *FEBS Lett.*, **197**, 315-320.

McCarthy, J.E.G. & Gualerzi, C. (1990). Translational control of prokaryotic gene expression. *Trends Genet.* **6**, 78-85.

Miller, J.H. (1972). Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 352-355.

Min Jou, W., Haegeman, G., Ysebaert, M. & Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* **237**, 82-88.

Olsthoorn, R.C.L. (1996) PhD thesis, Leiden University, Leiden, The Netherlands.

Pon, C. & Gualerzi, C.O (1986). Mechanism of translational initiation in prokaryotes. *FEBS Lett.* **195**, 215-219.

Poot, R.A., Tsareva, N.V., Boni, I.V. & van Duin, J. (1997) RNA folding kinetics regulates translation of phage MS2 maturation gene. *Proc.Natl. Acad. Sci USA* **94**, 10110-10115.

Renhof, R., Cielens, I., Nikitina, T., Sherina, L., Shomstein, Z. & Gren, E. (1985). Synthesis and functional activity of translation initiation regions in mRNA. *FEBS Lett.* **185**, 277-281.

Ringquist, S., Shinedling, S. Barrick, D., Green, L., Binkley, J., Stormo, G.D. & Gold, L. (1992). Translation initiation in *Escherichia coli*: Sequences within the ribosome-binding site. *Mol. Microbiol.* **6**, 1219-1229.

Ringquist, S.T., Jones, T., Snyder, E.E., Gibson, T., Boni I. and Gold, L. (1995). High-affinity RNA ligands to *Escherichia coli* ribosomes and ribosomal protein S1: comparision of natural and unnatural binding sites. *Biochemistry,* **34**, 3640-3648.

Robertson, H.D. & Lodish, H.F. (1970) Messenger characteristics of nascent bacteriophage RNA. *Proc. Natl. Acad. Sci. USA* **67**, 710-716.

Romero, A. & Garcia, P. (1991) Initiatiation of translation at AUC, AUA and AUU codons in Escherichia coli. *FEMS Microbiol. Lett.* **84**, 325-330.

Sprengart, M.L. and Porter A.G. (1997). Functional importance of RNA interactions in selection of translation initiation codons. *Mol.Microbiol.* 19-28.

Shine, J. & Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**, 1342-1346.

Skripkin, E.A., Adhin, M.R., de Smit, M.H. & van Duin, J. (1990). Secondary structure of the central region of bacteriophage MS2 RNA. *J.Mol.Biol.* **211**, 447-463.

Steitz, J.A. (1969). Polypeptide chain initiation: Nucleotide sequences of three ribosomal binding sites in bacteriophage R17 RNA. *Nature,* **21**, 957-964.

Stormo, G. (1987) Translational regulation in bacteriophages. In: Translational Regulation of Gene Expression (J. Ilan, eds) Plenum Publishing , New Yourk, pp.27-49. Subramanian, A.-R. (1983). Structure and functions of ribosomal protein S1. *Progr. Nucl. Acid Res.* **28**, 101-142.

Tzareva N.V., Makhno, V.I and Boni, I.V. (1994). Ribosome-messenger recognition in the absence of the Shine-Dalgrno interactions. *FEBS Lett.* **337**, 189-194.

Uhlenbeck, O.C., Carey, J., Romaniuk, P.J., Lowary, P.T. and Becket, D. (1983). Interaction of R17 coat protein with its RNA binding site for translational repression. *J. Biomol. Str. Dyn.* **1**, 539-552.

van Dieijen, G., van Knippenberg, P.H. & van Duin, J. (1976). The specific role of ribosomal protein S1 in the recognition of native phage RNA. *Eur. J. Biochem.* **64**, 511-518.

vanDuin, J. (1988) Single-stranded RNA bacteriophages. In The viruses (H. Fraenkel-Conrat & R.Wagner, ed.) pp.117-167, Plenum Press, New York.

van Himbergen, J., van Geffen, B. & van Duin. (1993). Translation control by a long range RNA-RNA interaction; a basepair substitution analysis. *Nucl. Acids Res.* **21**, 1713-1717.

Witherell, G.W., Gott, J.M. and Uhlenbeck, O.C. (1991). Specific interaction between RNA phage coat proteins and RNA. *Prog. Nucl. Acid Res. Mol. Biol.* **40**, 185-220.

Zinder, N.E. (ed.) (1975). *RNA phages*. Cold Harbor Laboratory, Cold Spring Harbor, NY.

APPENDIXES

NUCLEOSIDES & NUCLEOTIDES, 14(3-5), 1047-1048 (1995)

TEMPLATE ACTIVITY OF OLIGORIBONUCLEOTIDES SYNTHESIZED BY THE PHOSPHORAMIDITE METHOD USING 2'-O-TETRAHYDROFURANYL AND 2'-O-TETRAHYDROPYRANYL PROTECTING GROUPS

E.Stankevich, V.Kumpins, N.Licis, J.Klovins, V.Berzin
Biomedical Research and Study Centre, University of Latvia,
1 A.Kirhensteina Str., LV-1067, Riga, Latvia

Abstract. Solid-phase synthesis and functional activity of oligoribonucleotides containing native and modified translation initiation region (TIR) of phage MS2 and fr RNA replicase gene have been investigated.

Solid-phase synthesis of 15-24 membered oligoribonucleotides was carried out with a Pharmacia Gene Assembler. Tetrahydropyranyl (Thp) and tetrahydrofuranyl (Thf) groups were employed for the 2'-hydroxyl protecting. We combined them with the use of 5'-dimethoxytrityl (DMTr) groups. Nucleoside phosphoramidite units were prepared from 2'-O-tetrahydropyranyl (or tetrahydrofuranyl) suitable base-protected nucleoside derivatives using bis(diisopropylamino)methoxyphosphine. CPG-550 and Bio-Glas-500 served as solid supports for the synthesis of oligoribonucleotides. Till now the combination of the acid-labile 2'-OThp (or 2'-OThf) and 5'-ODMTr was recommended only for the synthesis of short oligomers due to their unstability during chain elongation.[1]

The longer oligoribonucleotides (21 mers) were obtained by means of phosphoramidite approach using the levulinyl group for the protection of 5'-OH group and Thf for 2'-OH.[2,3]

We performed the coupling in CH_3CN during 3 min in the presence of 1-H-tetrazole. Dichloroacetic acid (1%) in dichloromethane was used for the removal of the DMTr group.

After cleaving from the support and deprotection with conc. aqueous NH_3 oligomers were separated by FPLC (C_8 column) and after treating with 80% CH_3COOH on a C_{18} column. All synthetic templates were sequenced enzymatically.

The synthetic oligoribonucleotides were used to investigate the specific interaction between ribosome and TIR of the prokariotic mRNA. Based on the previous studies of template activity of short fragments of phage MS2 and fr RNA containing TIR of the replicase gene, the approximate structure of the initiation region has been determined.[4] The latter contains the initiation codon and a 5'-sequence from it not exceeding 17 nucleotides and containing the Shine-Dalgarno sequence. We also studied oligoribonucleotides containing native and modified TIR's of phage MS2 and fr replicase gene. Functionally active 20mer frR(-17 → 3) have been selected for the template modifications with nucleotide substitutions and deletion. This RNA fragment contains the initiator AUG, the native SD region and 7 adjacent 5'-terminal nucleotides (Table).

Nucleotides within this region were changed by enzymatic or chemical synthetic methods. Oligomers 1 and 2 with the base changes in the spacer region of TIR and at the 5'-terminus had high template activity. The shortening of a sequence at the 5'-end of frR(-17 → 3) up to 16 mer frR(-13 → 3) also has no influence on the effectivity of translation

TABLE. Template activity of phage fr RNA fragments containing TIR of replicase gene

Oligoribonucleotide	Presence of TIR elements structure				Relative activity, %
	Actual AUG	SD sequence	Out-of-frame AUG	Terminator UGA	
Sequence of TIRs in native:					
fr R(-17→3)	+	+	+	+	100
5'CAACAUGAGGAAUACCCAUG3'					
MS2 R(-17→3)	+	+	+	+	100
5'AAACAUGAGGAUUACCCAUG'3'					
Synthetic:					
1 AAACAUGAGGAAUACCCAUG	+	+	+	+	100
2 AAACAUGAGGAAAACCCAUG	+	+	+	+	95
3 AAACAUGAAGAAUACCCAUG	+	-	+	+	27
4 AAACAUGAAGAAAACCCAUG	+	-	+	+	27
5 AAACACGAGGAAAACCCAUG	+	+	-	-	11
6 **AACAUGAGGAAUACCCAUG**	+	+	+	+	100
7 **AUGAGGAAUACCCAUG**	+	+	+	+	100
8 **UGAGGAAUACCCAUG**	+	+	-	+	59
9 **UCACCAAUACCCAUG**	+	-	-	-	9

Ribosome binding activities of oligoribonucleotides were compared at 40 pmol of input template.

Actual initiation codon of replicase gene is bold, the putative SD sequence is underlined, sequence AUGA of overlapping out-of-frame AUG and terminator codons is marked with upper line.

Oligoribonucleotides 1-5 were products of enzymatic synthesis, templates 6-9 were obtained from solid-phase synthesis.

Nucleotide substitutions are in bold type.

initiation complex formation. Loosing of the next 5'-terminal nucleotide (oligonucleotide 8) causes a decrease in activity. Base changes (G → A) leading to the modification of SD region (oligomers 3 and 4) sharply decrease the template activity. Single base changes (U_{-12} to C_{-12}) completely inactivate the template. In the latter case the out-of-frame initiation codon AUG and terminator UGA in AUGA sequence close to the 5'-end of SD sequence were eliminated.

These results suggest that the 70S ribosomal initiation complex formed by template oligonucleotide with start codon and the native Shine-Dalgarno tetranucleotide AGGA can be enhanced by presence of the AUGA sequence of overlapping initiation and termination codons.

References

1. S.Iwai, E.Ohtsuka, Nucleic Acids Res., 1988, 16, 9443-9456.
2. S.Iwai, T.Sasaki, E.Ohtsuka, Tetrahedron, 1990, 46, 6673-6688.
3. R.Kierzek, M.H.Caruthers, C.E.Longfellow, D.Swinton, D.H.Turner, S.W.Freier, Biochemistry, 1986, 25, 7840-7846.
4. V.Berzin, I.Cielens, J.Jansone, E.Gren, Nucleic Acids Res., 1982, 10, 7763-7775.
5. R.Renhof, I.Cielens, T.Nikitina, L.Sherina, Z.Shomstein, E.Gren, FEBS Letters, 1985, 185, 277-281.

II

БИООРГАНИЧЕСКАЯ ХИМИЯ, 1995, том 21, № 12, с. 920 - 924

УДК 547.963.32.07

15 - 19-ЗВЕННЫЕ СИНТЕТИЧЕСКИЕ ОЛИГОРИБОНУКЛЕОТИДЫ КАК МОДЕЛИ ИНИЦИАТОРНОГО УЧАСТКА ТРАНСЛЯЦИИ ГЕНА РЕПЛИКАЗЫ ФАГА fr

© 1995 г. В. Х. Кумпиньш, Н. Г. Лицис, Р. Ф. Ренхоф, Э. И. Станкевич, В. М. Берзинь*

Исследовательский и учебный центр биомедицины, Университет Латвии, Латвия, LV 1067, Рига, ул. А. Кирхенштейна, 1

Поступила в редакцию 05.12.94 г.

Проведен автоматический твердофазный синтез олигорибонуклеотидов длиной 15 - 19 н. о. с использованием 5'-О-диметокситритил-2'-О-(2-тетрагидропиранил)- (или 2'-О-(2-тетрагидрофуранил))-N-ацилрибонуклеозид-3'-О-(метил-N,N-диизопропил)фосфамидитных синтонов и 1-Н-тетразола в качестве активирующего агента. Сравнительный анализ матричной активности синтезированных олигорибонуклеотидов – моделей инициаторного участка трансляции гена репликазы фагов MS2 и fr показал, что минимальным активным фрагментом РНК является 16-членный нуклеотид, содержащий инициаторный AUG-кодон гена, короткий спейсер и домен Шайна–Дальгарно, а также 5'-концевую последовательность AUGA с функционально важным терминаторным кодоном AUG.

Ключевые слова: 5'-О-диметокситритил-2'-О-(2-тетрагидрофуранил)- (или 2'-О-(2-тетрагидропиранил))-N-ацилрибонуклеозид-3'-О-(метил-N,N-диизопропил)фосфамидиты; олигонуклеотиды, химический синтез; мРНК-модели; фаговые РНК, матричная активность фрагментов.

Специфическое взаимодействие рибосомы с участком инициации трансляции на прокариотических мРНК является ключевой стадией, определяющей точность и эффективность считывания матрицы. Основные структурные элементы сайта связывания рибосомы – инициаторный кодон AUG (редко GUG или UUG), последовательность Шайна–Дальгарно, спейсер между упомянутыми детерминантами матрицы, а также определенная вторичная структура инициаторного участка [1, 2]. Однако результаты статистического анализа [3] и данные экспериментальных исследований функциональной активности ряда участков инициации трансляции [4 - 9] убедительно доказывают, что кроме вышеупомянутых детерминантов в процессе инициации участвуют и другие последовательности мРНК, расположенные рядом со стартовым кодоном. В наших предыдущих публикациях [4, 10] было показано, что минимальный функционально активный участок инициации трансляции гена репликазы фагов MS2 и fr длиной в 20 нуклеотидов содержит истинный кодон инициации и тетрануклеотид Шайна–Дальгарно.

В настоящей работе с целью более подробного изучения всех структурных детерминантов инициации был разработан и проведен химический синтез ряда олигорибонуклеотидов, содержащих нативные и целенаправленно измененные последовательности участка инициации, и изучена их активность в связывании рибосом (образовании рибосомного комплекса инициации).

Полученные данные свидетельствуют, что 70S-комплекс инициации эффективно образуется на 16-членной нуклеотидной минимальной матрице, которая кроме истинного AUG-кодона и последовательности Шайна–Дальгарно содержит также расположенный на 5'-конце тетрануклеотид AUGA, соответствующий перекрывающимся последовательностям кодонов инициации и терминации.

Твердофазный синтез матричных олигорибонуклеотидов

Существенные результаты в синтезе олигорибонуклеотидов в настоящее время достигнуты при наращивании цепи на твердой фазе фосфамидитным методом при сочетании *трет*-бутилдиметилсилильной защитной группы для 2'-ОН и монометокситритильной или диметокситритильной группы для 5'-гидроксила [11, 12]. Однако

Сокращения: Thf – 2-тетрагидрофуранил, Thp – 2-тетрагидропиранил, DMAP – диметиламинопиридин, TEAA – триэтиламмонийацетат.

* Автор для переписки.

выделение конечных продуктов осложнено (в отличие от метода с применением ацетальных 2'-О-защитных групп) необходимостью очистки целевого олигонуклеотида от тетрабутиламмоний-фторида, который используется для отщепления *трет*-бутилдиметилсилильной группы. Есть также данные о применении 5'-О-левулинильной защитной группы вместо групп тритильного типа в сочетании с 2'-О-ацетальной группой [13], что, однако, сильно осложняло сам олигорибонуклеотидный синтез. Применение нуклеозидов, содержащих замещенные 4-метоксипиперидильные группы в качестве 2'-О-защиты [14 - 16], также ограничено из-за многостадийного синтеза исходных соединений. Следует упомянуть о синтезе 6 - 19-звенных олигорибонуклеотидов Н-фосфонатным методом с использованием 5'-О-диметокситритил-2'-О-(2-тетрагидропиранил)-N-ацилрибонуклеозид-3'-Н-фосфонатов [17]. Мы выбрали сочетание кислотолабильных 5'-О-диметокситритильной и 2'-О-(2-тетрагидрофуранильной) (или 2-тетрагидропиранильной) групп в фосфамидитном синтезе. Выбор сделан вопреки утверждению [18 - 21], что применение этих защит не позволяет получать олигорибонуклеотиды длиной более 12 н. о. Для 5'-О-детритилирования во избежание частичного отщепления Thf- и Thp-групп мы применяли 1% $CHCl_2CO_2H$ в CH_2Cl_2 (1.5 мин) [17].

В качестве истинной матрицы с помощью ферментативных методов был синтезирован функционально активный 20-членный олигорибонуклеотид frR(–17–3) [10], содержащий стартовый кодон AUG гена репликазы, последовательность нативного спейсера и тетрануклеотид Шайна–Дальгарно, а также дополнительно семь 5'-концевых нуклеотидов (таблица).

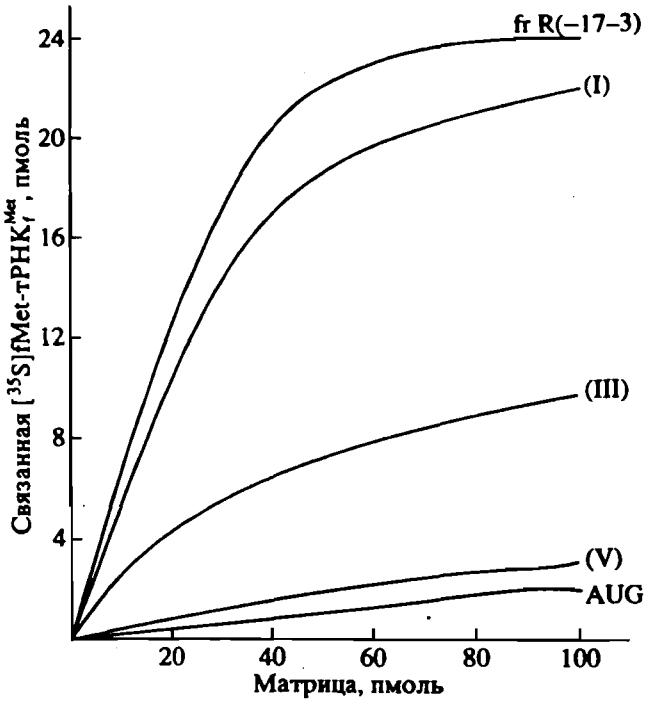
Далее при помощи методов химического синтеза [22 - 27], а в некоторых случаях энзиматического синтеза [10] был синтезирован также ряд олигорибонуклеотидов с различными изменениями в нуклеотидной последовательности матрицы. Чистоту и первичную структуру синтезированных олигонуклеотидов проверяли ПААГ-электрофорезом препаратов в денатурирующих условиях и энзиматическим секвенированием матрицы. Были синтезированы два аналога нативного frR(–17–3), где у одного олигорибонуклеотида, (I), 5'-концевой нуклеотид C_{-17} был заменен на A_{-17} (что наблюдается

Матричная активность фрагментов РНК фага fr, содержащих инициаторный участок трансляции гена репликазы*

Олигорибонуклеотид**	Присутствие элемента структуры участка инициации трансляции				Относительная активность, %
	инициаторный AUG	последовательность SD	ложный AUG	терминатор UGA	
Нативная матрица: frR (–17 –3)					
5' CAACAUGAGGAAUACCCAUG	+	+	+	+	100
Синтетические матрицы:					
(I) AAACAUGAGGAAUACCCAUG	+	+	+	+	100
(II) AAACAUGAGGAAAACCCAUG	+	+	+	+	95
(III) AAACAUGAAGAAUACCCAUG	+	–	+	+	27
(IV) AAACAUGAAGAAAACCCAUG	+	–	+	+	27
(V) AAACACGAGGAAAACCCAUG	+	+	–	–	11
(VI) AACAUGAGGAAUACCCAUG	+	+	+	+	100
(VII) AUGAGGAAUACCCAUG	+	+	+	+	100
(VIII) UGAGGAAUACCCAUG	+	+	–	+	59
(IX) UCACCAAUACCCAUG	+	–	–	–	9

* Состав реакционной смеси см. "Экспер. часть", количество матрицы 40 пмоль.

** Инициаторный кодон репликазы подчеркнут двумя чертами, последовательность Шайна–Дальгарно – одной чертой, последовательность AUGA, образующаяся от перекрывающихся ложного кодона инициации AUG и кодона терминации UGA, обозначена пунктиром. Олигорибонуклеотиды (I) - (V) являются продуктами энзиматического синтеза, матрицы (VI) - (IX) синтезированы при помощи химических методов. Нуклеотидные замены показаны жирным шрифтом.



Матричная активность синтетических моделей инициаторного участка трансляции гена репликазы фага fr, полученных с использованием олигорибонуклеотидов (I), (III), (V).

в инициаторном участке фага MS2 [4, 10]), а у другого, (II), была внесена дополнительная замена U_{-5} на A_{-5}. Далее с целью более точного установления роли 5'-концевой структуры участка инициации был получен ряд укороченных вариантов матрицы (олигонуклеотиды (VI) - (IX)).

Кроме того, были внесены замены, приводящие к разрушению района Шайна–Дальгарно (олигонуклеотиды (III), (IV)) и/или к удалению перекрывающихся кодонов инициации и терминации в блоке AUGA, расположенном на 5'-конце матрицы. Важно отметить, что все замены и делеции, введенные в последовательность исходного frR(–17–3), сохраняли или снижали стабильность вторичной структуры матриц.

Функциональная активность олигорибонуклеотидов

Для определения функциональной активности мы использовали метод фиксации комплекса инициации на мембранных фильтрах, измеряя количество инициаторной [^{35}S]fMet-тРНК$_f^{Met}$ на 70S-рибосомах *E. coli* в присутствии олигорибонуклеотидов в возрастающей концентрации (рисунок).

Как видно из таблицы, синтетические олигонуклеотиды (I) и (II), содержащие замены нуклеотидов в спейсерном районе и на 5'-конце инициаторного участка, демонстрируют высокую матричную активность, сравнимую с активностью исходного олигонуклеотида frR(–17–3). Удаление 5'-концевых звеньев, вплоть до 16-членного олигонуклеотида (VII) (frR(–13–3), содержащего инициаторный кодон AUG гена репликазы, спейсер и блок AUGA на 5'-конце), также не оказывает сколько-нибудь заметного влияния на эффективность образования инициаторного комплекса.

Потеря следующего 5'-концевого нуклеотида вызывает некоторое снижение способности полученного олигорибонуклеотида frR(–12–3) (матрица (VIII)) стимулировать связывание инициаторной [^{35}S]fMet-тРНК$_f^{Met}$ с рибосомой. Если учесть, что при укорачивании олигорибонуклеотида ликвидируется расположенный вне зоны считывания ложный инициаторный кодон AUG, то данные по наличию матричной активности у нуклеотида (VIII) указывают на отсутствие сколько-нибудь значительной роли этого триплета в связывании рибосом. Такое предположение подтверждают наши предыдущие данные [4] о "молчании" ложного AUG в составе более длинных фрагментов РНК инициаторного участка гена репликазы.

Олигонуклеотиды (III) и (IV) содержат специфичную нуклеотидную замену G_{-9} на A_{-9}, устраняющую природную последовательность Шайна–Дальгарно в составе исходного frR(–17–3). Такие матрицы проявляют заметно сниженную, но устойчивую активность (рисунок и таблица), что в какой-то мере подтверждает высказанное ранее предположение [28] о том, что комплементарные взаимодействия домена Шайна–Дальгарно с 3'-концевой последовательностью 16S рибосомальной РНК не играют исключительной роли в регулировании эффективности инициации трансляции. Более серьезные замены в этом районе (матрица (IX)) полностью устраняют инициаторную активность олигорибонуклеотида.

Матричная активность почти полностью теряется при точечной замене U_{-12} на C_{-12} (матрица (V)). В данном случае одиночная замена нуклеотида одновременно приводит к удалению как ложного инициаторного (AUG), так и терминаторного (UGA) кодона в блоке AUGA, расположенном на 5'-конце матрицы рядом с последовательностью Шайна–Дальгарно. Приведенные выше данные указывают на то, что кодон терминации трансляции UGA, по-видимому, является существенным элементом фрагмента инициаторного участка гена репликазы и участвует в процессе связывания рибосом. Ранее были высказаны предположения, что такие терминаторные кодоны, как UAA и UGA, составляют часть инициаторных участков трансляции мРНК и необходимы при подготовке рибосом перед инициацией [29, 30].

В итоге наши данные позволяют заключить, что минимальный активный в связывании рибосом 16-членный инициаторный участок гена

репликазы фагов fr и MS2 должен содержать истинный инициаторный кодон AUG, короткую последовательность спейсера, домен Шайна–Дальгарно, а также на 5'-конце от него последовательность AUGA с функционально важным терминаторным кодоном UAG.

Последовательность AUGA оказывает существенное влияние на процесс инициации трансляции in vitro, что в какой-то мере подтверждается количественными данными тоупринта рибосомных комплексов инициации на вариантах мРНК гена *rIIB* фага T4 [9], свидетельствующими о том, что короткие последовательности на 5'-конце от домена Шайна–Дальгарно существенны для эффективной инициации трансляции.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

В работе использовали пластины с PEI-целлюлозой (Merck, Германия), РНКазы T1 и U2 (Sankeo, Япония), РНКазу *Bacillus cereus*, нуклеазу P1, T4-РНК-лигазу и T4-полинуклеотидкиназу (Pharmacia, Швеция), ATP, GTP и тРНК$_f^{Met}$ (Boehringer-Mannheim, Германия), нитроцеллюлозные фильтры BA85 (Schleicher und Schuell, Германия), [^{32}P]ATP и [^{35}S]Met ("Изотоп", Россия).

Химический синтез олигорибонуклеотидов проводили на автоматическом синтезаторе Gene Assembler (Pharmacia, Швеция). Использовались 5'-O-диметокситритил-2'-O-(2-тетрагидрофуранил)- или 2'-O-(2-тетрагидропиранил)-N-ацилрибонуклеозиды (Biolar, Латвия), ацетонитрил ("Реахим", Украина), дихлоруксусная кислота, 1H-тетразол, коллидин, DMAP (Fluka, Швейцария), молекулярные сита 3Å (Ferak, Германия). Метиленхлорид перегоняли над P_2O_5 и CaH_2, ацетонитрил – над P_2O_5, а затем дважды над CaH_2 и выдерживали 15 ч над молекулярными ситами. Тетразол дважды сублимировали в вакууме. В качестве твердофазного носителя использовали стекло аминопропил-CPG-550 (Fluka, Швеция) и Bio-Glas-500 (Bio-Rad, США). Bio-Glas-500 аминопропилировали по методу [22]. Модификацию твердой фазы и присоединение первого нуклеозидного звена проводили согласно работам [23, 24]. Фосфамидиты синтезированы по описанным в литературе методам [18, 25, 26]. В основном использовалась схема операции для серийного синтезатора Gene Assembler. Конденсацию проводили в присутствии 1H-тетразола в течение 5 мин. Колонку фирмы Pharmacia, из которой нами был удален твердый стандартный носитель, наполняли 20 мг аминопропил-CPG-550 с присоединенным первым нуклеозидом. Количество введенного нуклеозида составляло в среднем 40 - 60 мкмоль/г.

Удаление защитных групп и выделение олигорибонуклеотидов. После завершения последней стадии конденсации к носителю добавляли 600 мкл конц. NH_4OH (*d* 0.89) и выдерживали 16 ч при 55°C. Раствор декантировали, а носитель 2 раза промывали водно-этанольным (1 : 1) раствором. После упаривания сухой остаток растворяли в 0.1 M TEAA (pH 7.1) и центрифугировали. Супернатант наносили на колонку C_{18}-PepRPC HR 5/5 (Pharmacia, Швеция) и проводили хроматографию в обращенной фазе (ОФХ) в системе быстрой хроматографии белков FPLC (Pharmacia, Швеция) в градиенте концентрации ацетонитрила (0 - 40%) в 0.1 M TEAA. Выделяли фракцию, содержащую целевые 5'-O-диметокситритил-2'-O-(2-тетрагидрофуранил)- или 5'-O-диметокситритил-2'-O-(2-тетрагидропиранил)-защищенные олигонуклеотиды. Далее эту фракцию обрабатывали 80% уксусной кислотой в течение 3 ч при 37°C. Образец упаривали досуха, растворяли в 0.1 M TEAA и выделяли при помощи ОФХ (см. выше). Скорость элюции 1 мл/мин.

Индивидуальность синтезированных олигорибонуклеотидов проверяли электрофоретически в 18% ПААГ в присутствии 7 M мочевины; их нуклеотидная последовательность была подтверждена гель-электрофорезом продуктов энзиматического гидролиза [31].

Ферментативный синтез олигорибонуклеотидов осуществляли как описано ранее [10]. Нуклеотидную последовательность продуктов проверяли двумерным фракционированием P1-нуклеазных гидролизатов: первое направление – высоковольтный электрофорез на ацетате целлюлозы (pH 3.5), второе – хроматография на PEI-целлюлозе при 65°C.

Стимулируемое олигорибонуклеотидами связывание [^{35}S]fMet-тРНК$_f^{Met}$ с рибосомами. Реакционная смесь (50 мкл) содержала 0.05 M трис-HCl (pH 7.8), 0.1 M NH_4Cl, 5 мМ $MgCl_2$, 0.001 M дитиотреит, 0.002 M GTP, 60 пмоль 70S рибосом, очищенные факторы инициации: IF1 – 0.4, IF2 – 3.77 и IF3 – 0.63 мкг, 60 пмоль [^{35}S]fMet-тРНК$_f^{Met}$ синтетические олигорибонуклеотиды (20 - 100 пмоль). После инкубации при 37°C в течение 10 мин смесь фильтровали через мембранные фильтры, промывали и количество связанной [^{35}S]fMet-тРНК$_f^{Met}$ определяли в жидкостном сцинтилляционном спектрометре (Intertechnique SL4000, Франция).

СПИСОК ЛИТЕРАТУРЫ

1. *Stormo G.D.* // Maximizing Gene Expression / Eds W. Reznikoff, L. Gold. Boston: Butterworth, 1986. P. 195 - 224.
2. *de Smith M.H., van Duin J.* // Progr. Nucl. Acids Res. Mol. Biol. 1990. V. 4. P. 1 - 35.
3. *Schneider T.D., Stormo G.D., Gold L.* // J. Mol. Biol. 1986. V. 188. P. 415 - 431.
4. *Berzin V., Cielens I., Jansone I., Gren E.* // Nucl. Acids Res. 1982. V. 10. P. 7763 - 7775.

5. *Bingham A.M., Busby S.J.M.* // Mol. Microbiol. 1987. V. 1. P. 117 - 124.
6. *Olins P.O., Rangwala S.H.* // J. Biol. Chem. 1989. V. 264. P. 16973 - 16976.
7. *Deyfus M.* // J. Mol. Biol. 1988. V. 204. P. 79 - 84.
8. *Boni I.V., Isaeva D.M., Musychenko M.L., Tzareva N.V.* // Nucl. Acids Res. 1991. V. 19. P. 155 - 162.
9. *Hartz D., McPheeters D.S., Gold L.* // J. Mol. Biol. 1991. V. 218. P. 83 - 97.
10. *Renhof R., Cielens I., Nikitina T., Sherina L., Shomstein Z., Gren E.* // FEBS Lett. 1985. V. 185. P. 277 - 281.
11. *Ogilvie K.K., Usman N., Nicoghosian K., Cedergren R.I.* // Proc. Natl. Acad. Sci. USA. 1988. V. 85. P. 5764 - 5768.
12. *Lyttle H.M., Wright P.B., Sinha N.D., Bain J.D., Chamberlin A.R.* // J. Org. Chem. 1991. V. 56. P. 4608 - 4615.
13. *Iwai S., Sasaki T., Ohtsuka E.* // Tetrahedron. 1990. V. 46. P. 6673 - 6688.
14. *Beijer B., Sulston I., Sproat B.S., Rider P., Lamond A.I., Neuner P.* // Nucl. Acids Res. 1990. V. 18. P. 5143 - 5151.
15. *Rao M.V., Reese C.B., Schehlmann V., Yu P.S.* // J. Chem. Soc. Perkin Trans. 1. 1993. P. 43 - 55.
16. *Sakatsume O., Ohtsuki M., Takaku H., Reese C.B.* // Nucl. Acids Res. 1989. V. 17. P. 3689 - 3697.
17. *Веньяминова А.Г., Горн В.В., Зенкова М.А., Комарова Н.И., Репкова М.Н.* // Биоорган. химия. 1990. Т. 16. С. 941 - 950.
18. *Tanaka T., Fujino K., Tamatsukuri S., Ikehara M.* // Chem. Pharm. Bull. 1986. V. 34. P. 4126 - 4132.
19. *Ohtsuka E., Yamane A., Ikehara M.* // Chem. Pharm. Bull. 1983. V. 31. P. 1534 - 1543.
20. *Kierzek R., Caruthers M.H., Longfellow C.E., Swinton D., Turner D.H., Freier S.M.* // Biochemistry. 1986. V. 25. P. 7840 - 7846.
21. *Iwai S., Ohtsuka E.* // Nucl. Acids Res. 1988. V. 16. P. 9443 - 9456.
22. Oligonucleotide Synthesis. A Practical Approach / Ed. M.I. Gait. Oxford: IRL Press, 1984. P. 46.
23. *Ефимов В.А., Бурякова А.А., Ревердатто С.В., Чахмачева О.Г.* // Биоорган. химия. 1983. Т. 9. С. 1367 - 1387.
24. *Ломакин А.И., Ястребов С.О., Попов С.Г.* // Биоорган. химия. 1985. Т. 11. С. 920 - 926.
25. *Веньяминова А.Г., Косолапова З.А., Левина А.С.* // Биоорган. химия. 1987. Т. 13. С. 125 - 127.
26. *Barone F.D., Tang I.Y., Caruther M.N.* // Nucl. Acids Res. 1984. V. 12. P. 4051 - 4061.
27. *Никитина Т., Шомштейн З., Шерина Л., Ренхоф Р., Грен Э.* // Биоорган. химия. 1986. Т. 12. С. 482 - 489.
28. *Cologero R.A., Pon C.L., Canomeco M.A., Gualerzi C.O.* // Proc. Natl. Acad. Sci. USA. 1988. V. 85. P. 6427 - 6431.
29. *Ganoza M.C.* // Can. J. Biochem. 1977. V. 55. P. 267 - 281.
30. *Atkins J.F.* // Nucl. Acids Res. 1977. V. 7. P. 1035 - 1041.
31. *Donis-Keller H., Maxam A.M., Gilbert W.* // Nucl. Acids Res. 1977. V. 4. P. 2527 - 2538.

Synthetic 15 - 19-meric Oligoribonucleotides as Models for Translation Initiation of the Phage fr Replicase Gene

V. Kh. Kumpin'sh, N. G. Litsis, R. F. Renhof, E. I. Stankevich, and V. M. Berzin'[1]

Research and Education Center of Biomedicine, Latvian University, ul. A. Kirchensteina 1, Riga, LV 1067 Latvia

Abstract – Automatic solid-phase synthesis of 15 - 19-meric oligoribonucleotides was carried out using 5'-*O*-dimethoxytrityl-2'-*O*-(2-tetrahydropyranyl)- [or 2'-*O*-(2-tetrahydrofuranyl)-] *N*-acylribonucleoside-3'-*O*-(methyl-*N*,*N*-diisopropyl)phosphoramidite synthons and 1-H-tetrazole as an activator. Comparative analysis of the template activity of the oligoribonucleotides synthesized, which are models of the translation initiation region of the replicase gene of MS2 and fr phages, showed that the minimal active fragment of RNA is a 16-mer containing the initiation AUG codon of the gene, a short spacer, a Shine–Dalgarno domain, and the 5'-terminal AUGA sequence with a functionally important termination AUG codon.

Key words: oligoribonucleotide phosphoramidites; oligonucleotides, chemical synthesis; mRNA models; phage RNA, template activity of enzymes.

[1] To whom correspondence should be addressed.

III

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

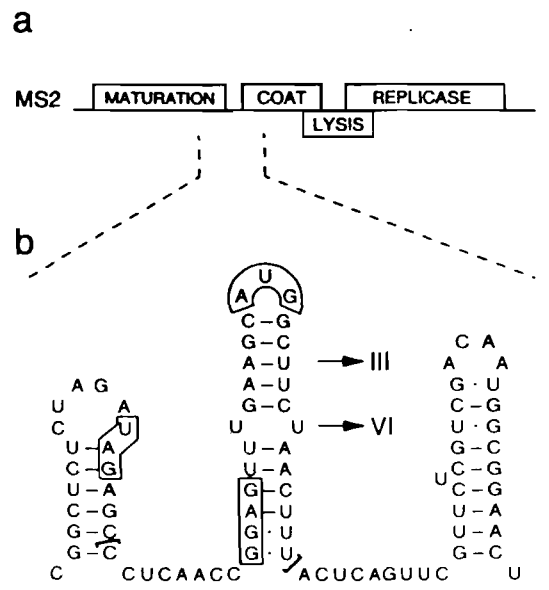


Fig. 1. (a) Map of the MS2 RNA genome. (b) Secondary structure at the start of the coat protein gene. Its initiation codon, the Shine–Dalgarno sequence and the UAG stop codon of the poorly translated maturation protein gene are boxed. Roman numerals are used to number the base-pairs in the stem. Brackets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

replicase genes (van Himbergen *et al.*, 1993). In Q_β RNA the binding sites for Q_β 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	5×10^{11}	7×10^{10}	3×10^{8}	2×10^{7}	4×10^{10}	7×10^{6}
42°C	2×10^{11}	2×10^{11}	3×10^{8}	1×10^{8}	1×10^{11}	9×10^{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_β (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

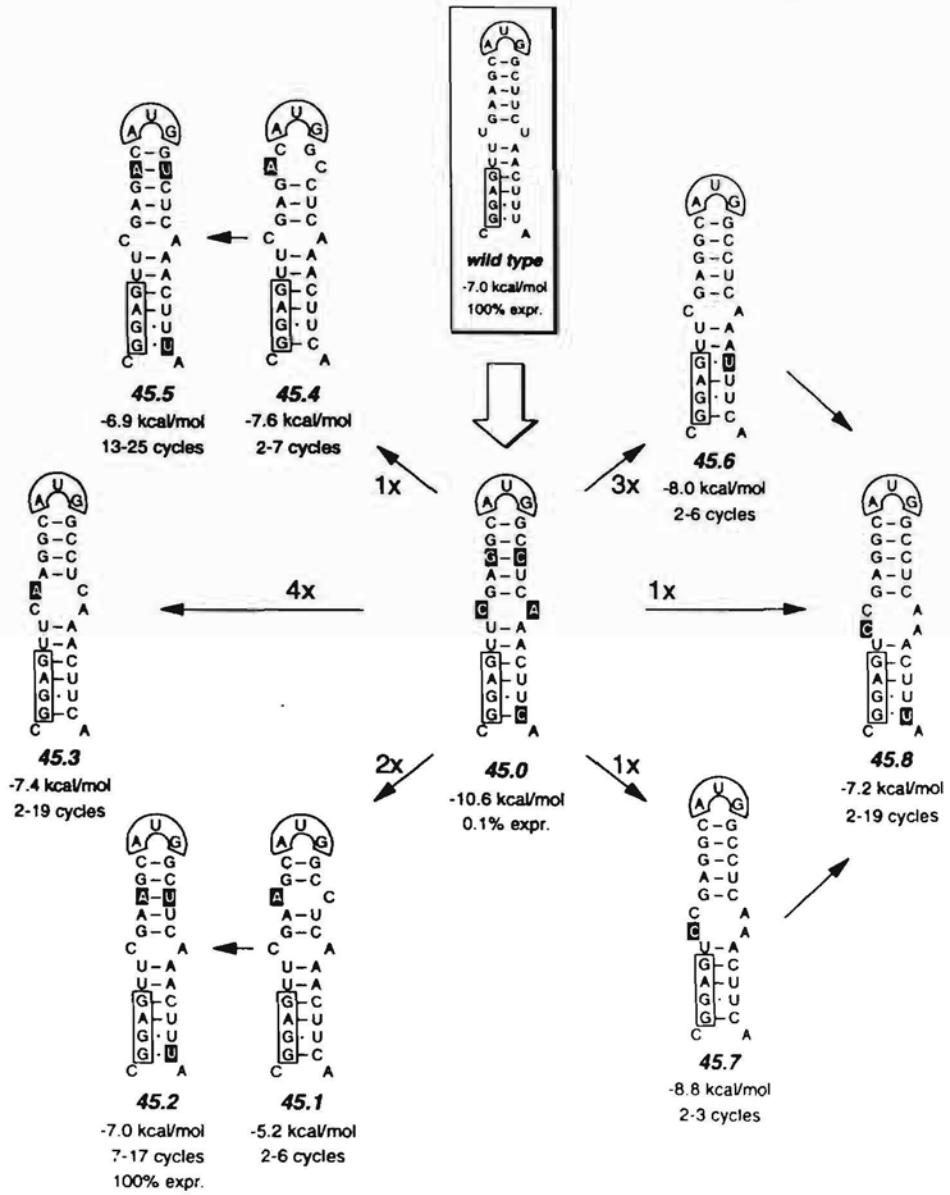


Fig. 2. Evolutionary pathways leading from non-viable mutant 45.0 to various viable species. Numbers at arrows pointing to revertants show how many plaques were found with this sequence. The wild-type structure is presented for convenience. Black boxes in the mutant sequence mark the differences from 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

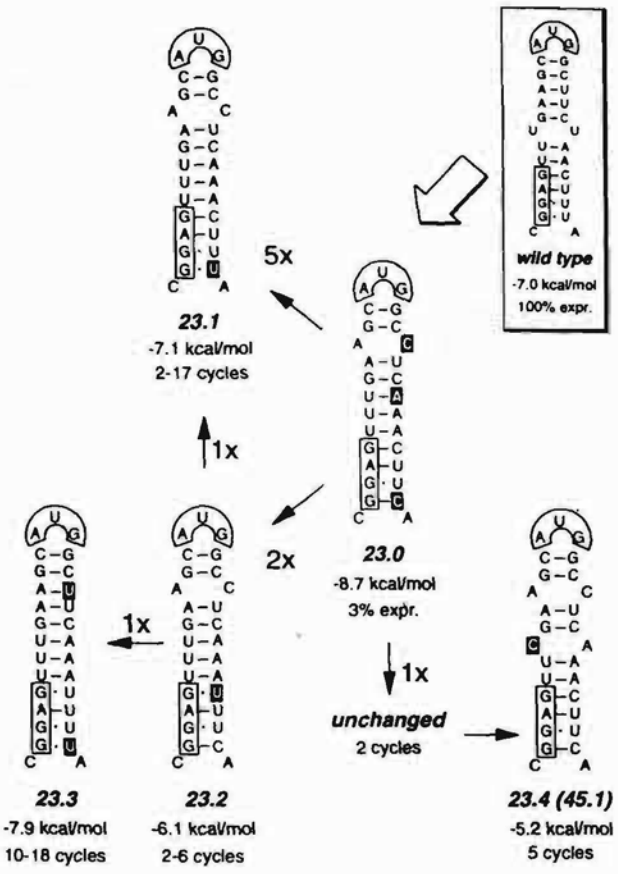


Fig. 3. Revertants of mutant 23.0. Revertant 23.4 is identical to 45.1. See legend to Figure 2 for further details.

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.

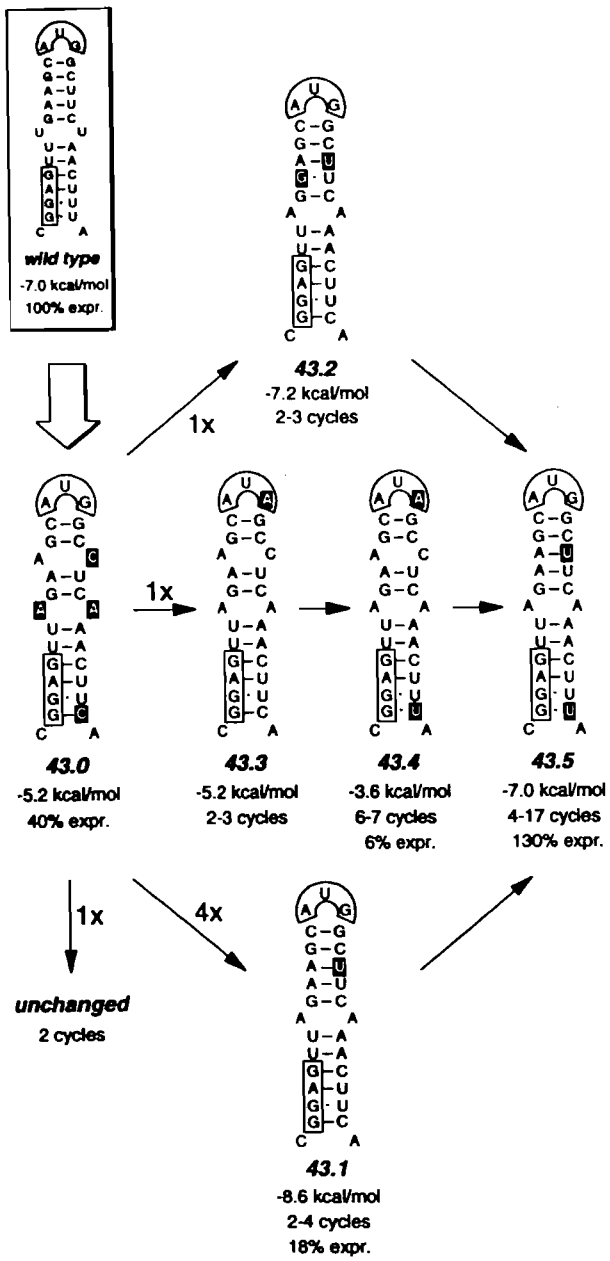


Fig. 4. Revertants of mutant 43.0. See legend to Figure 2 for further details.

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 $\Delta G°$ 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 $\Delta G°$ 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 $\Delta G°$, 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 $\Delta G°$ 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 $\Delta G°$; 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 $\Delta G°$ 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

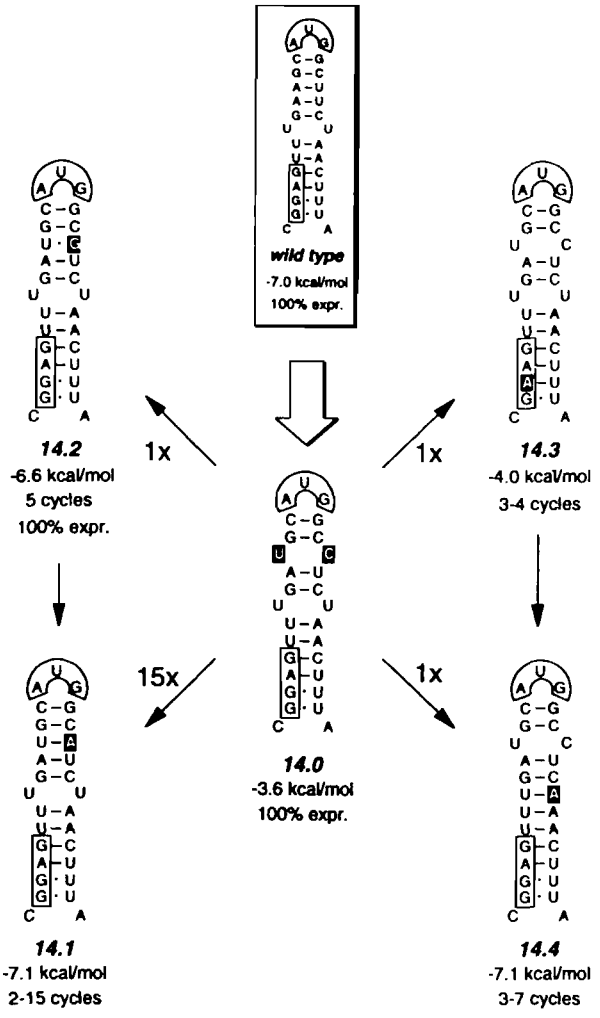


Fig. 5. Revertants of mutant 14.0. See legend to Figure 2 for further details.

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

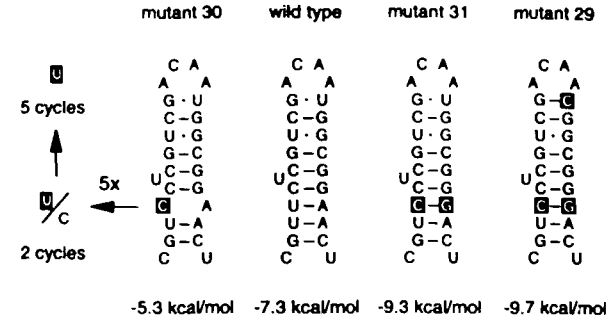


Fig. 6. Revertants of mutants in the second coat gene helix. Mutants 31 and 29 were unchanged after 13 cycles.

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

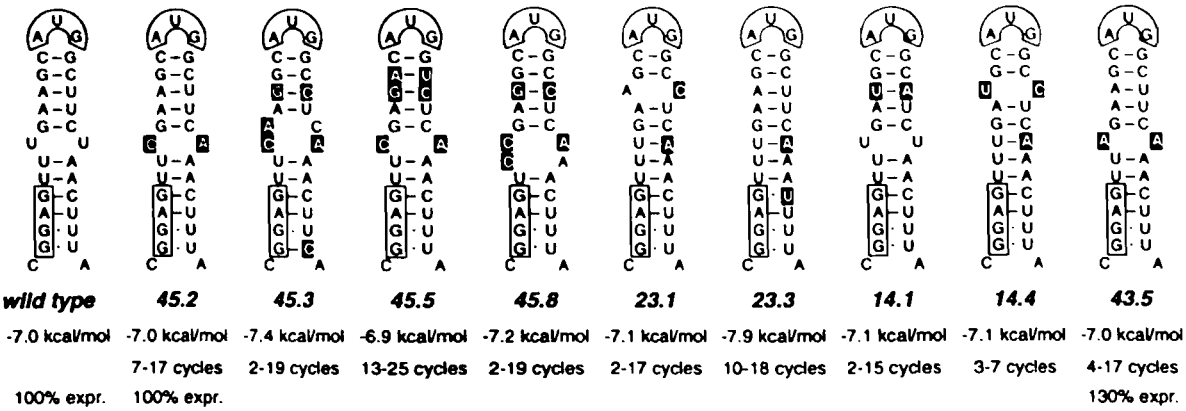


Fig. 7. Compilation of coat initiator helices in stable revertants.

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 ~ 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_β RNA this question has been studied *in vitro* and the common site, the S site, overlaps the start region of the Q_β 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 $:rpA_{am}$, $lacZ_{am}$, $Sm^r/\lambda dbio_{252}$, $cI_{857}\Delta H_1$), 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 *Xba*I(1303)–*Bst*XI(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.
Bont,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

IV

Evolution of translation initiation region of the phage MS2 RNA replicase gene from scratch

Normunds Licis, Zita Balklava and Valdis Berzins*
Biomedical Research and Study Centre, University of Latvia, Ratsupites 1, LV1067, Riga, Latvia
*Corresponding author

SUMMARY

The translation initiation region (TIR) of the replicase gene of RNA bacteriophage MS2 contains a conventional ribosome-binding site (RBS) defined by an AUG codon and a SD-sequence. In addition to initiation of replicase translation this region also takes part in negative control over ribosome binding to the replicase gene start. The RBS adopts a specific stem-loop structure, the so-called translational operator, and serves as a high affinity binding site for the phage coat protein that acts as a translational repressor. By site-directed mutagenesis we randomized an essential sequence of the replicase initiation site in an infectious MS2 cDNA clone. The obtained phage-producing clones lacked a native-like RBS and produced replicase more than 100-times less efficiently than wild type. The introduced mutations lowered the phage titer by two and more orders of magnitude. Nine of the selected mutants were allowed to replicate in *E.coli* F^+ bacteria to monitor evolutionary improvement of the replicase gene TIR. In course of adaptation the mutants acquired 1-5 suppressor base changes upstream the translation initiation AUG codon and restored a high phage viability. The general evolutionary pattern was creation of various SD-sequences in the replicase initiation site and we observed no other common sequence-specific features. In addition, one phage had rebuilt an RNA operator hairpin but number of others evolved partially resembling structures. We suggest that these resembling helixes can serve as low affinity phage RNA coat protein binding sites in course of infection. We also found that the level of the replicase gene translation in phage RNA pseudorevertants is high as compared to parental clones but low as compared to wild type and in some instances are exceedingly dependent on the upstream coat gene sequences. These findings are discussed with respect to the control of the replicase gene translation in phage MS2.

INTRODUCTION

The process of translation initiation in prokaryotes determines the fidelity and efficiency of gene expression (Gold, 1988; McCarthy and Gualerzi, 1990; de Smit and van Duin, 1990 a, b). Three kinds of interactions are known to take place during selection of translation initiation region (TIR) on prokaryotic mRNAs. First, ribosomal protein S1 is involved in mRNA binding (van Dieijen *et al.*, 1978; Hartz *et al.*, 1990). It seems to have a preference for U-rich regions (Zhang and Deutsher, 1992; Boni *et al.*, 1992) but may also recognize sequence or structure specific motifs (Tzareva et al., 1994; Ringquist et al., 1995). Second, interaction between 3'-end of 16S rRNA and poly-purine stretch, the so-called Shine-Dalgarno (SD) sequence preceding the translational start codons in bacterial messages, provide the ribosomes with affinity for TIR (Shine and Dalgarno, 1974; de Smit and van Duin, 1994).

This SD-interaction is conceivable for the initiation on most of bacterial messengers and is the only one rRNA-mRNA interaction that has gained direct experimental support (Hui and de Boer, 1987). Other sequences within RBS have been proposed to interact with 16S rRNA but they either lack an evidence in recognition of a TIR through rRNA-mRNA interactions or the experimentally data are contradictory (reviewed by McCarthy and Gualerzi, 1990; Sprengart and Porter, 1997). At least some of these so-called translational enhancers may serve as binding sites for S1 protein (Tzareva et al., 1994; Ringquist et al., 1995). Finally, the interaction between mRNA initiation codon and fMet-tRNA anticodon turns the ribosome-mRNA complex into virtually irreversible state (McCarthy and Gualerzi, 1990). Initiation may use all, two or only the codon-anticodon interaction. Apart from the positive determinants of the initiation process the accessibility of a ribosome-binding site (RBS) on mRNA can be controlled in a negative way by secondary structure of the TIR. Strong intramolecular interactions may turn a good translational start into virtually inaccessible for ribosomes (de Smit and van Duin, 1990 a, b) and probably is a mean to hide false initiation sites within coding regions (Ganoza *et al.*, 1987).

Despite extensive studies uncertainty still surrounds the essential determinants of a prokaryotic TIR. Thus, some of the *in vitro* data (Ravetch and Jakes, 1976; Calogero *et al.*, 1988; Melancon *et al.*, 1990) seem to stand under question the relevance of SD-interaction for the translation initiation. Also, many analyses (Gold, 1988; Dreyfus, 1988 and references cited therein) imply that within the RBS there are other sequence specific properties, *e.g.* nucleotide content, for their distinction from non-initiator regions. We reasoned that the question on significance of certain mRNA TIR properties for translation initiation can be assessed exploiting natural evolution system available for the single stranded RNA bacteriophage MS2. The mutations in phage RNA can be introduced through its full-length cDNA copy present on a plasmid and evolution of the phage mutants to high viability by means of compensatory mutations can be monitored (Olsthoorn *et al.*, 1994). Thus if a RBS of a phage gene is substituted with unrelated sequences such mutant RNA regions are likely to become adapted for an appropriate efficiency of translation initiation again.

Phage MS2 RNA genome (Fig.1) codes for four proteins. One of them, the replicase protein, associates with several host proteins and serves for phage RNA replication. Since the phage development is dependent on supply of replicase we expect no or low phage titer if the RBS of the replicase gene is unsatisfactory for an adequate replicase translation and an increase in phage titer upon improvement of the initiation site.

The replicase gene TIR is defined by a SD-box present upstream of the initiator AUG codon (Berkhout, 1986; Stankevich *et al.*, 1995). A sequence that involves the SD-region was randomized in infectious MS2 cDNA clone and the evolutionary improvement of the TIR was monitored upon phage passaging.

Additional reason to choose this particular TIR as the target for analysis lied in a rather sophisticated control of the replicase translation in wild type phage. Although the replicase gene RBS is efficient itself, dual negative control limits replicase synthesis to only a small burst early in the infection (reviewed by van Duin, 1988). One control is provided by long-distance interaction (LDI) of an internal coat gene segment to the nucleotides surrounding the replicase gene start. LDI restricts the initiation of the replicase gene translation unless it is broken up by ribosomes reading the coat message. The major molecular basis for this control is the MJ-interaction (van Himbergen *et al.*, 1993) but a nearby 1419-1426/1810-1817 LDI possibly contribute to the replicase suppression (Licis *et al.*, 1998). Another regulatory mechanism operates later during infection. When the phage coat protein accumulates in the cell it binds to the

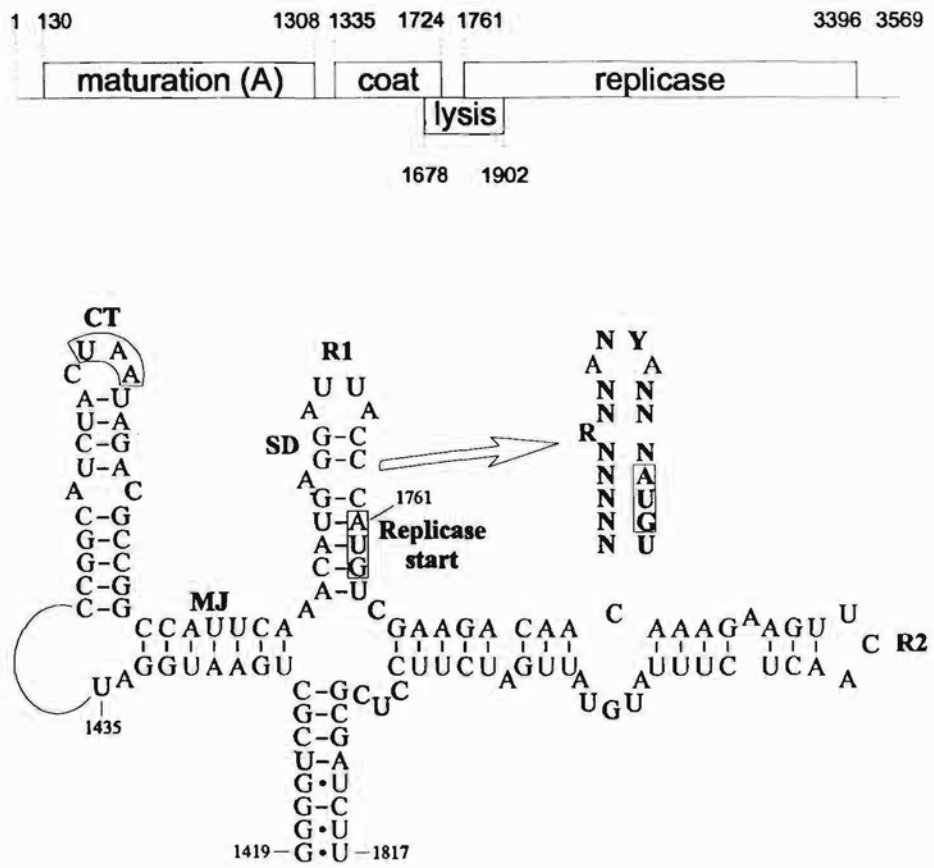


Figure 1. Genetic map of RNA coliphage MS2 and a model for the RNA secondary structure of the replicase gene TIR. Local stem-loop structures are indicated as CT, R1, R2. The start codon of the replicase and the stop codon of the coat gene are marked by box. SD – Shine Dalgarno sequence, MJ designates the MJ-interaction. The mutations were introduced in the indicated sequence of the replicase gene TIR.

replicase initiator hairpin, the so-called translational operator, and thereby blocks ribosome access to the replicase gene start. The features necessary for this protein-RNA interaction to take place are well known (Uhlenbeck *et al.*, 1983; Witherell *et al.*, 1991; Schneider *et al.*, 1992; Peabody, 1997; see at Fig.1). The RNA secondary structure is important for coat protein binding but the sequence generally is not. The exceptions are requirements for nucleotide specificity at the bulge that must be purine and three out of four hairpin loop positions. The RNA operator-coat protein interaction was thought to mediate the phage RNA encapsidation (Beckett and Uhlenbeck, 1988) but was shown recently to be unessential for that purpose (Peabody, 1997). Because the mutations in the replicase gene start region must destroy the native hairpin it was of interest how the phage will adapt to the lack of translational repression.

In addition, the replicase gene TIR is a part of the phage lysis (L) cistron that is translated in +1 frame with respect to replicase. The lysis aminoacids encoded in the replicase initiation site are non-essential for the function of L protein (Berkhout *et al.*, 1985) and the region either plays no role in the control of the lysis gene expression (Scmidt *et al.*, 1987). Since that the RNA sequence constraints implied by the L gene on the evolution of the replicase gene TIR are absent or at a low level, except for the cases when mutations prevent the translation of the entire L gene (non-sense codons *etc*).

Table 1. Oligonucleotides used in this study

primer	sequence	position; strand	application
RP57[a]	GAACTTCTTTGTTGTCTTCGACATNNNTRNTNNYNNNN NTTGAATGGCCGGCGTCTA	1728-1784; -	mutagenesis
p114	CATCCGGATCCCATGACAAGG	2047-2067; -	RT, PCR
p162	TTGAATGGCCGGCGTCTA	1728-1745; -	PCR
p9	CGAAGACAACAAAGAAGTTC	1765-1784; +	PCR
p180	TCGAGGGGTACAATCCGT	1200-1217; +	RT, PCR
p108	GAATTCCGACTGCGAGCTTAT	1628-1648; +	sequencing
p106[b]	GGTAAATTTCAAGAGAAAGATCG	1811-1833; -	RT, sequencing
p127	GCCCCATAGTGGCACCG	2153-2169; -	RT, PCR

RT-reverse transcription

[a] N-any nucleotide, R-purine, Y-pyrimidine

[b] underlined residue mark mismatch

MATERIALS AND METHODS

Bacterial strains and plasmids

Plasmids were grown in *E.coli* K12 strain M5219, encoding the thermosensitive λ repressor (cI857) and the transcriptional antitermination factor N (Remaut *et al.*, 1981). Evolution experiments were performed in the *E.coli* F^+ strain AB259 (Hfr3000, Thi^-, Su^-). All strains were grown on LB broth.

pMS2000 harbors a full-length DNA copy of phage MS2 RNA under transcriptional control of the thermoinducible promoter P_L from phage λ (Olsthoorn *et al.*, 1994). This plasmid produces phage spontaneously even without induction of the promoter and after overnight growth at 28°C the supernatant of bacterial cultures contains about 10^{11} pfu/ml. Plasmid pD (Licis *et al.*, 1998) differs from pMS2000 by that it contains a short linker substituting the 1303-1901 MS2 cDNA XbaI-BfrI restriction fragment and thereby does not yield the phage. Plasmid pD was used to reconstruct full-length MS2 cDNA by ligation of the PCR-generated MS2 cDNA mutant fragments into the XbaI and BfrI restriction sites of the vector.

Plasmids used for the measurements of the replicase gene expression contained either 1365-2057 or 1628-2057 MS2 cDNA fragment behind the promoter P_L, and lac Z gene fused at the BamHI site to the replicase gene (nt 2057). The wild type construct is MJL0 and contains the 1365- 2057 MS2 cDNA fragment (van Himbergen *et al.*, 1993). Other clones were constructed by substituting it with a mutant 1365-2057 fragment. pCOAT184 contains the phage coat protein gene under control of constitutive *tet* promoter (Berkhout, 1986). In this construct *tet* promoter ensures a low level of coat protein sufficient to repress replicase gene translation by about 90%. The plasmid confers to cells chloramphenicol resistance.

Mutagenesis and cloning

Oligonucleotide RP57 was used to introduce mutations in 1746-1760 region of MS2 genome. It contains randomized target TIR sequence flanked by native phage MS2 genome sequences. At first, the flanking MS2 cDNA fragments were amplified in PCR 1 and PCR 2 reactions using praimer pairs p114-p9 and p162-p180, respectively (Fig.2). These fragments were gel-purified, precipitated and dissolved in 50 µl of water. In the mutagenesis step (PCR3) the reaction mixture (40 µl) was combined from 5 µl of 10x PCR buffer (100 mM TRIS (pH 8.8), 500 mM KCl, 0.8% NonidetP40), 4 µl 25 mM $MgCl_2$, 1 µl BSA (1mg/ml), 8 µl of 1 mM dNTPs, 5 µl of 5 µM p180, 5 µl of PCR2 product, 2 µl of 0.5 µM oligonucleotide RP57 and 2 units of Taq DNA polymerase. After 5 cycles (denaturation at 95°C for 30 sec, annealing at 45°C for 40 sec and elongation at 72°C for 40 sec) 5 µl of PCR1 product and 5 µl of primer p114 were added and the reaction was continued for 25 more cycles. The PCR product was precipitated, cDNA dissolved in water and after treatment with XbaI and BfrI restriction endonucleases ligated to the corresponding restriction sites of pD plasmid.

Recording phage evolution

To produce mutant phage M5219 F^- cells containing the mutant plasmids were grown

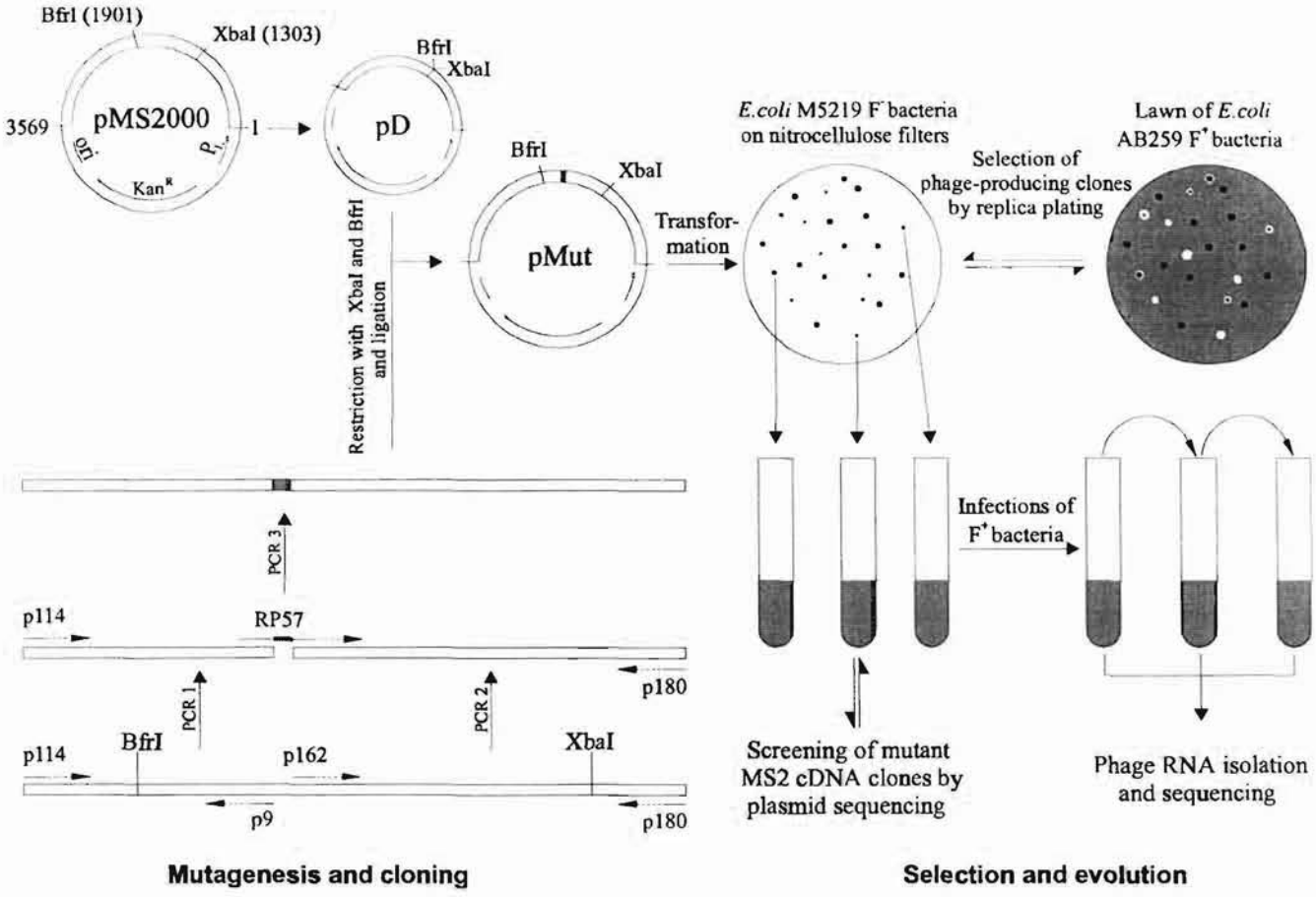


Figure 2. Outline of experimental approach. See Methods for the details.

overnight (cycle 1). 10^5-10^7 pfu of the produced phage were used to infect *E.coli* AB259 F^+ bacteria (at OD_{670} of about 0.2) in 2.5 ml of liquid LB broth and cultures were shaken overnight at 37°C (cycle 2). 10^5 pfu from the previous culture were taken for subsequent cycle and so on. The infections were continued up to cycle 20. The phage RNA was isolated from 1 ml of supernatant of the cultures (Olsthoorn *et al.*, 1994) and reverse transcribed with either primer p114 or p127. The RT-product was PCR-amplified with p180 and one of the above primers and the amplified cDNA was sequenced with T7 DNA polymerase, routinely using oligo p108 as described previously (Licis *et al.*, 1998). The phages for sequence analysis were taken after cycles 3, 10 and 20. If the examination of the RNA sequence suggested genetic heterogeneity in the phage population, plating was sometimes used to separate individual genomes for subsequent sequence analysis.

Probing RNA structure with S_1 nuclease

2 μg of phage RNA was preincubated in the reaction buffer (40 mM sodium acetate (pH 4.6), 340 mM NaCl, 1.4 mM $ZnCl_2$ and 7% glycerol) for 10 min at 37°C to equilibrate the secondary structure of the template. Subsequently, the sample was treated with 1 unit of S_1 nuclease for 10 min at 37°C. After phenol extraction and precipitation of the RNA, primer p106 was extended in 6 μl of 50 mM TRIS-HCl, pH 8.3, 4 mM KCl, 8 mM $MgCl_2$ and 10 mM DTT containing 1unit of AMV reverse transcriptase.

Assay of β-galactosidase

0.05 ml of fresh overnight cultures of bacteria harboring plasmids that contain lacZ gene fused to

Table 2. Properties of MS2 cDNA clones

Clone	Nucleotide sequence*	Titer, pfu/ml **
pMS2000 (Wt)	acau**gagg**auuacccaug	10^{11}
Random pool	NNNNNRNNANYANNNAUG	-
A2	CcUCAaAgaCCaUcGaug	10^{8}
A8	GAauCGCAaCCaGAGaug	10^{8}
A14	aGUuUGgCaACaGccaug	10^{9}
A16	aGauCagCauuacccaug	10^{8}
A21	UcUGUaUCaCCaAcAaug	10^{9}
A23	GAUACaCCaCuaUAGaug	10^{9}
A28	UcUACagCauuacccaug	10^{9}
A36	CAACUGCgaAuaUGcaug	10^{7}
A37	CcUuAGUUaCCaAccaug	10^{8}
Clones containing termination codon(s) in frame of L gene		
A1	CACugaAgauCacUGaug	10^{5}
A5	aACAUaUUaGCaccGaug	10^{5}
A7	GACAAaUAaGuaUUGaug	10^{6}
A35	UUGAAGUAaACaUGcaug	10^{6}
A41	CAGCUagCaACaUGcaug	10^{5}
A13	aAUuCGUCauuaGUAaug	10^{4}
A19	CAUuCGCUaAuaGAcaug	10^{4}
A20	UAaCUaCCaGuaGUAaug	10^{4}

*The 1746-1763 RNA sequence of the mutants is aligned with that of the wild type. Capital letters indicate the mutations. The replicase start codon and SD-sequence are in bold. The stop codons of the lysis gene sequence are underlined.

**Appropriate dilutions of supernatant from cDNA cultures grown overnight at 28°C were tested on lawns of *E.coli* AB259 F^+ cells.

the phage replicase gene or the above clones provided in addition with pCOAT plasmid were added to 5 ml fresh LB broth plus ampicillin or ampicillin and chloramphenicol, respectively. The cells were grown at 28°C until OD_{670} of about 0.25 was reached and then shifted to 42°C for induction of the P_L promoter, grown for additional 30 min with vigorous shaking, after which the cultures were placed on ice for at least 10 min. 0.4 ml of each culture was added to 0.6 ml of Z buffer (Miller, 1972), 40 µl of $CHCl_3$ and 20 µl of 0.1% SDS in an eppendorf tube. The rest of bacterial cultures were left on ice until the cell densities at OD_{570} were determined. The blank tube was made in the same manner but contained broth and ampicillin instead of bacterial culture. The tubes were vortexed for 10 sec and put at 28°C for 10min. Afterwards 0.2 ml of ONPG (4mg/ml in Z buffer) was added to each tube and they were incubated at 28°C for 10 – 100 min. The reactions were terminated by the addition of 0.5 ml of 1M Na_2C0_3 and the tubes were centrifuged briefly. The breakdown of ONPG was measured at OD_{410}. All measurements were done on microtiter plates. Because the wavelengths used to calculate the β-galactosidase activity (A_{570} for the cell densities and A_{410} for the breakdown of ONPG) do not correspond to the wavelengths used in routine calculation (Miller, 1972), the values are only approximately similar to Miller units of β-galactosidase activity.

RESULTS

The experimental strategy

The undertaken approach is depicted in Figure 2. The 1746-1760 sequence of the replicase gene TIR was randomized by PCR-directed mutagenesis on a cDNA fragment. The mutated fragment then was cloned into the pD plasmid to give rise the full-length

MS2 cDNA and the ligated DNA was transformed into the *E.coli* F^- host bacteria. The transformed cells were seeded onto nitrocellulose filters and grown to colonies. The filters were then replica plated onto the lawn of *E.coli* F^+ bacteria to identify those clones that produced the viable phage. A number of the phage-producing colonies were taken from the filters and grown in liquid broth for the MS2 cDNA sequence analysis. Subsequently, the phages were subjected to consecutive infections of F^+ bacteria to allow evolutionary improvement of the replicase gene TIR. The target sequence was monitored after various number of passages. Finally, some of the evolved pseudorevertants were analyzed with respect to efficiency and control of the replicase gene translation after cloning in an expression vector.

Because it was uncertain initially whether the phages lacking a high affinity coat protein binding site in the replicase TIR may be able to encapsidate their RNA the nucleotide specificity at certain positions was preserved to increase the chance for the infective virion formation.

Properties of mutants

About 20% of transformants produced the phage and the sequence of the replicase initiation site was determined for some 20 clones (Tab.2). They contained 3-11 base changes as compared to the wild type replicase gene TIR and lacked a conventional RBS (SD-sequence) and an RNA operator hairpin. Apart from these common properties, the constructs fall in two distinct categories with respect to the phage lysis gene (Tab.2). About half of the mutants possessed an unperturbed L gene frame and their titers were by 2-3 orders of magnitude lower than the titer of the wild type clone. Many others contained one or even two nonsense codons in frame of the L gene. These prevent the synthesis of the functionally essential C-terminal peptide of the lysis protein and present an additional, the coding defect. In those clones the phage titer had dropped more drastically. To avoid the influence of the coding demands on the evolution of the sequence and structure of replicase initiation site the lysis-deficient mutants were omitted from the present analysis. Note, however, that as revealed by later experiments their viable progeny phages possess an intact frame of the lysis gene restored in various ways.

Evolution

The evolution of phage RNA pseudorevertants was allowed to proceed for 20 cycles during which the phage titers increased to nearly wild type level showing that the improvement of the replicase gene TIR had occurred. In this course of adaptation the mutants acquired 1-5 nucleotide substitutions in the replicase initiation site (Fig.3) and those containing many compensatory base changes generally reached the highest viability. Typically, we did not encounter an original mutant after amplification in *E.coli* F^+ host, except for mutant A28. This phage preserved the unchanged sequence and a low titer up to cycle 3.

In some instances we monitored only one evolutionary pathway whereas in other cases various pseudorevertants were encountered (Fig.3). Illustrative example is a complex evolution of mutant A16 during that three different progeny phages - A16.1, A16.2 and A16.3 were found. The pseudorevertant A16.1 was the dominant specimen at the third cycle when the two other descendants A16.2 and A16.3 were not detectable yet, probably because they constituted only a very minor fraction of the phage population. Later after cycle 10, however, better revertants A16.2 and A16.3 outgrew descendant A16.1. Subsequently, at cycle 20 progeny phage A16.2 had disappeared but A16.3 evolved to A16.4.

The replicase translation initiation region of phage RNA pseudorevertants

(i) TIR sequence

The sequence data unambiguously revealed a common source of the improvement of the replicase start. It

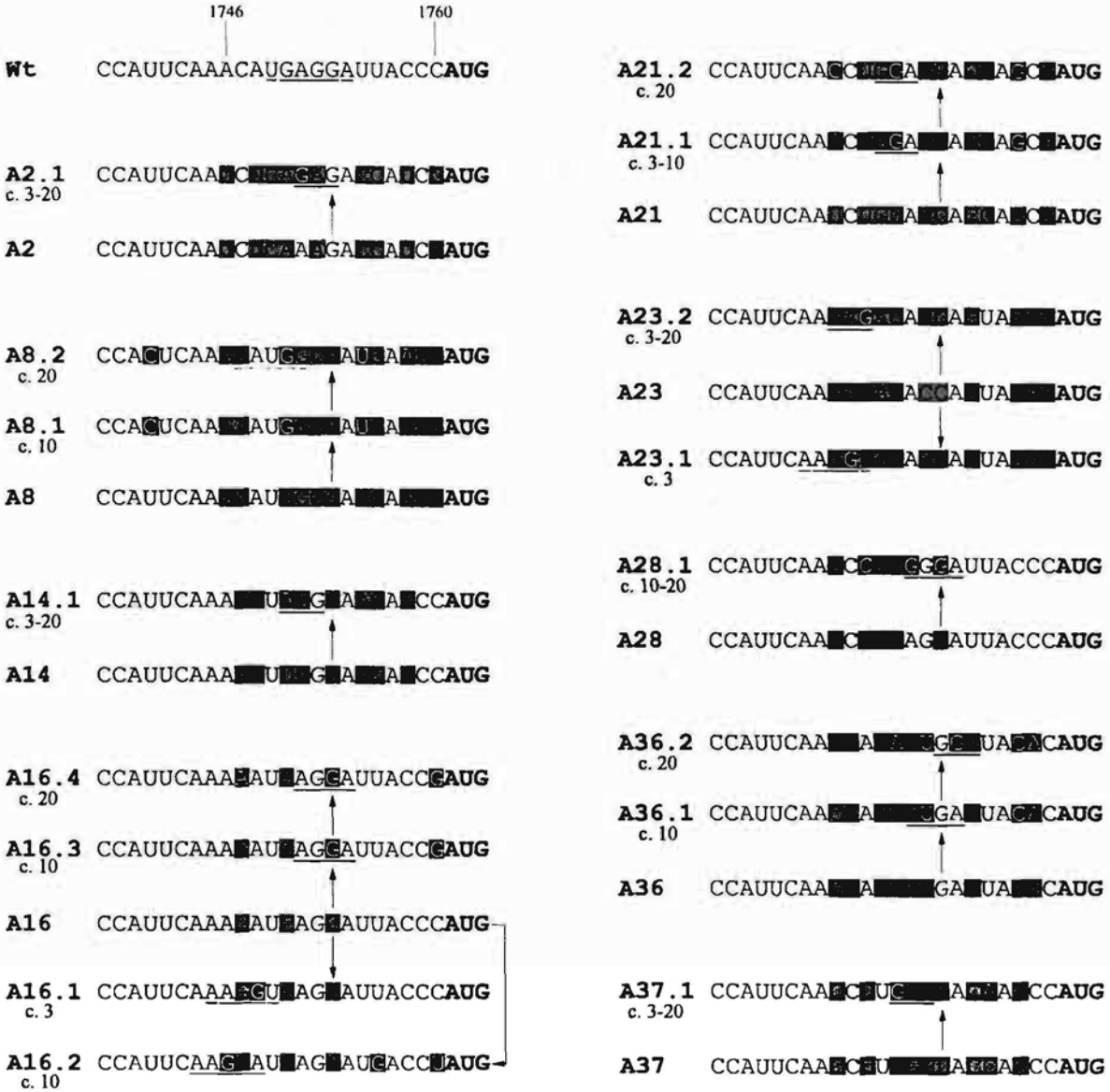


Figure 3. Sequence alignment showing the phage RNA evolution. Initial mutations are in gray while suppressor mutations in black boxes. The designation of pseudorevertants depicts their origin. For example, the progeny phages A16.1, A16.2, A16.3 derived from the mutant A16. 'c' stands for cycle. The putative SD-sequences in revertant RNA are underlined. The RNA sequences of the replicase TIR of revertants from A8 and A36 were poorly defined after the third cycle and therefor are not shown, while mutant A28 retained the original sequence after the third cycle.

appeared that various phages had evolved various SD-sequences (Fig.3 and Fig.4). This common event was mostly accomplished by base changes creating guanines and there seems to be only one exception from the SD-evolution. Revertant A8.2 does not posses a sequence to basepair with the 3'-end of 16S rRNA unless the mismatches in the SD-interaction are allowed (the region of a potential SD-box is indicated by broken line in Fig.3).

The restore of the SD-sequence was the only evolutionary event in the initiation site of several pseudorevertants (A2.1, A14.1, A16.1, A23.1 and A23.2) while other phage RNA acquired also one or more additional

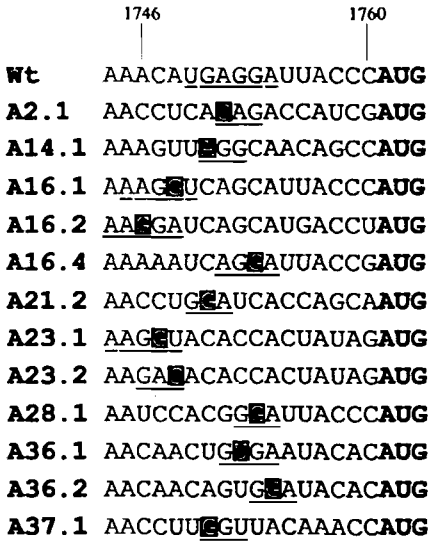


Figure 4. Sequence alignment showing the evolved SD-regions. The SD-related suppressor mutations are black boxed.

nucleotide substitutions. These base changes neither result in a common sequence specific pattern nor also resemble each other. Thus, descendant A8.2 selected the G1758A substitution located three nucleotides upstream the start AUG codon. In this instance the compensatory mutation might be interpreted as a phage attempt to increase the efficiency of the replicase translation. The data on randomized ribosome binding sites (Barrick *et al.*, 1994) revealed high activity of adenine at position –3 from the initiation codon that is consistent with the statistics for that position (Gold *et al.*, 1981). However, revertant A21.2 acquired just opposite A1758G base change showing that SD-unrelated compensatory mutations are specific for any given phage.

In wild type phage the RNA secondary structure of the replicase gene TIR is important for translational control by repression and coupling and thereby apart from the sequence specificity the secondary structure it is another conceivable target of the suppressor mutations. An illustrative example is the particular revertant A8.2 in which one of the suppressor mutations was located upstream the target sequence and in structure context (Fig.1) it brakes a basepair in the middle of the MJ-interaction. As pointed out before, the MJ-interaction suppresses the replicase translation and its disruption in wild type phage RNA activates the replicase synthesis (van Himbergen *et al.*, 1993; Licis *et al.*, 1998).

(ii) Structure of the replicase gene TIR

Computer program MFold (Zuker, 1989) that predicts both optimal and suboptimal RNA foldings indicated two alternative conformations for RNA of each revertant. One conformation is a wild type-like structures characterized by the presence of a replicase initiator hairpin (Fig.5) whereas the other does not contain a replicase initiator hairpin because a sequence of or around the translational start AUG codon is basepaired to the coat protein gene (not shown).

In those phage RNA that had evolved many base changes the compensatory substitutions seem to be targeted on creation of certain hairpins (Fig.5). Thus, two out of three base changes revertant A16.4 RNA appear to recreate a wild type-like translational operator. The two substitutions restoring the hairpin probably took place simultaneously (see evolution of this pseudorevertant at Fig.2). Two, also simultaneous, base changes stabilize an initiator hairpin in phage A21.2 RNA. In this case the resulting helix is not a high activity coat protein binding site. It, however, possesses the AUCA tetraloop that fulfils a requirement for coat protein-RNA interaction. Note, that the helix is displaced upstream the location of the replicase translational operator in wild type RNA and also does not contain the AUG start codon within the stem. Other examples of restoration of a replicase initiator hairpin are descendants A28.1 and A16.2. In revertant A28.1 the three compensatory base changes close a helix that contains the native loop but has a mismatch in the middle. In

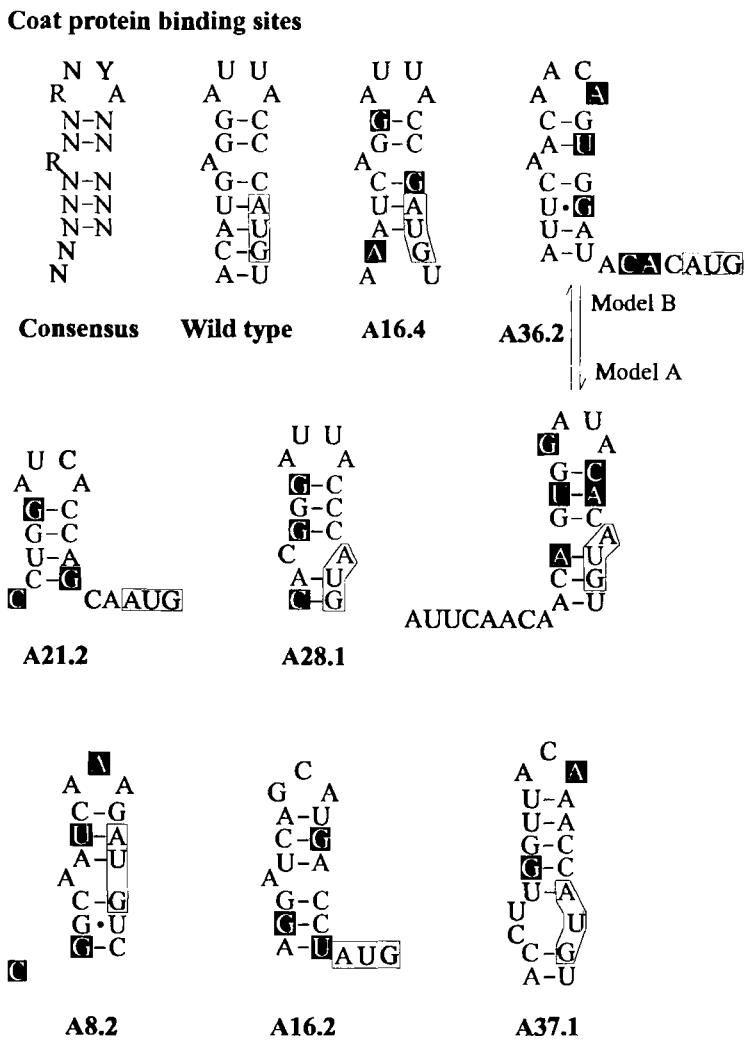


Figure 5. Possible secondary structure models for the replicase gene TIR in pseudorevertants A8.2, A16.2, A16.4, A21.2, A28.1, A37.1 and A36.2. Suppressor mutations are shown in black boxes. Wild type hairpin and the RNA structural requirements for binding binding by MS2 coat protein are shown for comparision. N represents any nucleotide, R and Y are purine and pyrimidine, respectivelly.

descendant A16.2 the three compensatory mutations appear to stabilize a hairpin with a loop of three nucleotides and a bulged A in the stem. Similar hairpin can be drawn for the phage A8.2 RNA as well. Also, a hairpin with loop of three nucleotides is predicted for the revertant A37.1 RNA. This phage, however, is uninformative from structural viewpoint because the only suppressor mutation stabilizing the stem of the hairpin is the SD-related and therefore its significance may lie solely in SD-sequence. For the same reason the secondary structure models for other revertant RNA are not presented.

Two possible stem-loop structures are shown for the progeny phage A36.2 RNA. In one model a replicase initiator hairpin is present in the original location (model A). The other folding may contain a high affinity coat protein binding site (found by eye, model

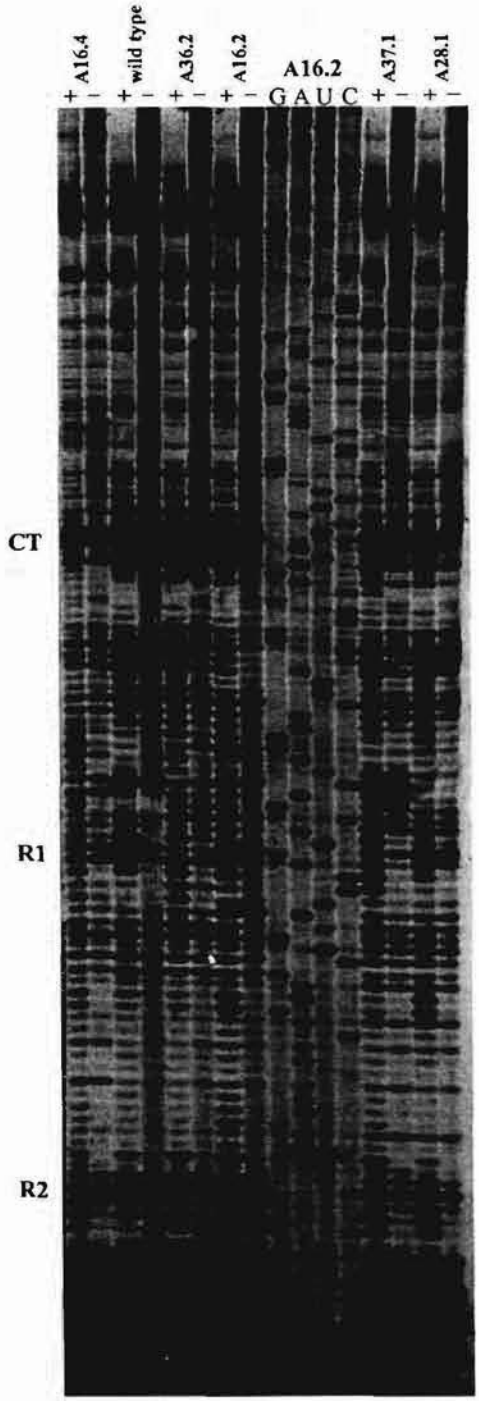


Figure 6. Nuclease S_1 structure probing of phage RNA from wild type and revertants A16.4, A36.2, A16.2, A37.1 and A28.1. The positions of the native replicase initiator hairpin R1 loop as well as of R2 and CT hairpin loops are indicated in the margin. Sites of cleavage were determined by primer extension. '+' designates S_1 digestion line, '-' control line. Revertant A16.2 was used for the sequencing lines.

B) but as suggested by the computer program the top of the hairpin is probably unfolded and forms a loop of nine rather than four nucleotides (not shown). An indication that the stem-loop might exist in the drawn form is provided by the suppressor mutation that turns AACU sequence present in progenitor RNA into AACA. Such a loop satisfies a requirement for coat protein binding site.

The structure models presented for A36.2 also illustrate the two alternative RNA conformations that as pointed out above might exist in revertant RNA. In the model B for the phage A36.2 RNA, the replicase initiation AUG codon and the surrounding nucleotides are displaced out of hairpin and they are accessible to basepair somewhere else, for instance with coat protein gene.

To get insight in the reliability of the structure models, we probed various pseudorevertant RNA with S_1 nuclease (exemplified at Fig.6), an enzyme that preferentially cuts at RNA hairpin loops. In overall, the structure probing data are in agreement with the presented helixes. In case of revertant A36.2 the results are consistent with both A and B models, indicating that the shown structures can exist at equilibrium. The data also suggest that the structure of the replicase gene TIR is poorly defined for all of the tested descendants that would be consistent with the existence of alternate structures.

(iii) Translational efficiency and control of the replicase gene

An important question is what is the efficiency of the replicase gene translation in pseudorevertants? Most of them lack a high affinity binding site for the coat protein and it means that efficient repression of the replicase gene in pseudorevertant RNA cannot take place. To assess the point cDNA of several pseudorevertant RNA, mostly those containing many suppressor mutations which are likely best adapted synthesis, as well as some of the parental mutants were cloned in an expression vector. In the used clones the replicase production can be measured as the

Table 3. Efficiency of replicase gene translation, translational repression and translational coupling in wild type, mutants and revertants.

RlacZ	Activity, β-galactosidase units[1]	Relative activity (%)[2]	Repression[3]	Coupling[4]
wild type	442	100	8.9	9
A8	2	<1	n.d.	20
A8.2	75	17	1.0 (1.0)	17
A16	1	<1	n.d.	57
A16.1	12	3	n.d.	n.d.
A16.2	13	3	1.0 (1.1)	95
A16.4	36	8	2.1	8
A21	2	<1	n.d.	3
A21.2	9	2	1.2 (1.5)	121
A36.2	38	9	1.1 (1.0)	35

RlacZ – MS2 cDNA clone in which the MS2 replicase gene is fused to the lacZ gene.

[1]The β-galactosidase activity for the clones containing the 1365-2057 cDNA sequence of MS2 genome.

[2]The wild type RlacZ activity is set at 100%.

[3]The repression is defined as the fold-decrease of the β-galactosidase activity of an RlacZ clone when the wild type coat protein is provided *in trans* from another plasmid. The values given in brackets correspond to the repression measured for the clones containing the 1628-2057 MS2 cDNA.

[4]The coupling correspond to the fold-increase of the β-galactosidase activity in RlacZ clones containing the 1628-2057 sequence of MS2 genome as compared to the corresponding clones that contain the1365-2057 sequence of MS2 genome.

activity of replicase-β-galactosidase (RlacZ) fusion protein. The phage cDNA in these constructs begin downstream the start of the phage coat protein gene (at position 1365 of MS2 genome) and thus coat gene translation is absent. However, since the clones contain most of the coat gene sequences, the replicase suppression by long-distance interaction(s), if such are present, may take place. Thereby the constructs mimic the native situation where the replicase initiation site is under the control of the coat protein gene. In addition, whether or not the replicase repression by wild type MS2 coat protein operates was measured by providing the coat protein *in trans* from pCOAT plasmid. This approach was intended to verify the reliability of the secondary structure models with respect to presence or absence of a translational operator hairpin.

The data (Tab.3) showed that the efficiency of the replicase gene translation in pseudorevertants is high as compared to the parental clones but low as compared to the wild type. The repression of the replicase synthesis by coat protein was detected only in descendant A16.4 which possesses a native-like hairpin, though less efficient than in wild type clone. Probably, the operator hairpin in phage A16.4 RNA is mostly unfolded.

These results suggest that the phage does not tend to maximize the replicase gene translation and, with respect to coat protein binding, are in agreement with the proposed structures. Also, it is clear that revertant A36.2 RNA does not possess a well distinct RNA operator hairpin.

The structural data indicated alternative revertant RNA conformation in which the replicase gene start is basepaired to the coat protein gene and the MJ-interaction (Fig.1) does not exist (see the models A and B for the

phage A36.2 RNA, Fig.5). If that has some true the dependence of the replicase production on the presence or absence of the coat gene sequences in pseudorevertants may differ from that in wild type clone. This can be deduced by removal of the coat gene sequences from the experimental constructs and re-determining the efficiency of the replicase synthesis. For this reason we deleted the MS2 cDNA from our constructs up to nucleotide 1628. Such a deletion removes most of the coat gene and in wild type clone activates the replicase translation about 9 times (Berkhout and van Duin, 1985). The increase is defined here as the coupling value. We found that only revertant A16.4 exerts a native-like coupling value. The efficiency of the replicase gene translation in other assessed descendants are characterized by an exceeding dependence on the presence of the 1365-1628 coat gene sequence and in two instances (A16.2 and A21.2) appeared to be even 10 times increased. In overall, these results are consistent with our recent data on the phage MS2 RNA mutants deficient in translational coupling. In that study we introduced the mutations in the MJ-interaction and some of the pseudorevertants acquired compensatory base changes either in replicase initiator hairpin or in the R2 helix.

Remarkably, here the suppression of the replicase by coat sequences is high even in revertant A8.2 in which RNA the MJ-interaction is at least weakened by a compensatory mutation. Thus, some RNA features other than MJ-interaction account for the suppression of the replicase translation in the revertants. Such RNA properties are most probably different LDIs between the coat and the replicase genes.

The increased coupling values rose the question whether the exceeding suppression of the replicase synthesis is established by evolution, for instance, to compensate for the lack of an efficient repression by coat protein binding. To assess the point the coupling value was estimated also for three mutants A8, A16 and A21. We detected increased coupling for the first two clones. Although this result did not exclude that the exceeding replicase gene suppression had been a target for the compensatory substitutions, it shows that the coupling value strongly depends on the sequence or structure of the replicase gene TIR and here may have arisen by chance.

The increased coupling values and the structure modeling of pseudorevertant RNA imply the (co)existence of alternate revertant RNA conformations. Since that the folding of the RlacZ transcript may differ significantly between clones starting at nt 1365 and 1628 of the MS2 genome and the replicase gene repression for some revertant RNA was reassessed also in the last constructs. In the 'shortened' clone the translation of A21.2 RNA seems to be somewhat repressed by wild type coat protein. This phage RNA contains the AUCA hairpin loop known to be of tighter binding phenotype (Lowary and Uhlenbeck, 1987). Likely, it may ensure some level of coat protein binding even in the absence of a bulged adenine.

DISCUSSION

In this study we have substituted an essential part of the naturally occurring phage MS2 replicase gene TIR with random sequences in an infectious MS2 cDNA clone. After identification by a selection procedure about 20 of the phage-producing clones were randomly picked up for further sequence analysis. Neither of the selected mutants contained a native-like replicase RBS - a SD-box and a translational operator hairpin. The mutants exerted a phage titer by two and more orders of magnitude lower than the titer of wild type. The clones yielding the least phage were found to contain nonsense codons in frame of the overlapping lysis gene and were omitted from the present analysis. Other phages were subjected to further studies and some of them were examined for the efficiency of the replicase gene translation. As expected, the mutations drastically decreased the replicase production and turned the replicase gene into translationally almost

silent. The mutants were further allowed to improve the replicase RBS by accumulation of the compensatory mutations upon passaging in *E.coli* F^+ bacteria. The essential features of evolution and its outcome were as follows. (i) Evolution of the RBS from false to true was accomplished almost invariably by creation of various SD-sequences. (ii) There seems to be a preference for specific RNA hairpins in the replicase gene TIR, even so they are not high affinity binding sites for the coat protein. (iii) The efficiency of the replicase gene translation in pseudorevertants is relatively low as compared to wild type.

Evolution of the SD-sequence

The SD-sequence appeared to be the general target of evolution and emerged in our study as the main positive determinant of a TIR. This becomes clear from several lines. First, various SD-sequences were evolved almost invariably in the progeny phage genomes (Fig.3 and Fig.4). Second, the SD-related mutations were the only ones selected in several pseudorevertant RNA.

The reconstruction of SD-sequence appeared to be very easy for the phage because it required only one appropriate base change nearby guanines present in parental RNA. This situation is best illustrated by the progeny of mutant A16 (Fig.3 and 4). Two of the descendants exploited the G present in AAAGAU region and turned the sequence in either AAGGAU (A16.1) or AAAGGU (A16.2). Still another progeny phage A16.3 used the AGCA region and exchanged it for AGGA SD-box. These unambiguous patterns are in line with well-known role of SD-sequence in translation initiation (Gold, 1988; Dreyfus, 1988; de Smit and van Duin, 1994). In fact, if the significance of a polypurine stretch were not known it could be deduced from our data.

Once created the SD-sequences were not subjected to further evolution despite that they mostly consisted of only three nucleotides (GAG, AGG, GGA and GGU) and could be extended by appropriate adjacent substitutions on both sides. Although there is no simple quantitative relationship between the strength of SD-sequence and the efficiency of translation initiation (de Smit and van Duin, 1994) this general pattern of non-evolution gives some indication that the phage does not tend to maximize the replicase synthesis. In one case, however, there seems to be selection for a different, probably more appropriate, SD-box. Phage A36.1 possessed a potential GUGA SD-sequence whereas its later development - phage A36.2 evolved GGA SD-box nearby.

Besides the SD-evolution no more common sequence-specific features were found to be characteristic for the various descendant RNA. This imply that other sequence specific RNA features within a prokaryotic RBS are not of a general relevance for translation initiation. On the other hand the initial mutant MS2 cDNA clones did not contain a SD-sequence but still yielded the phage. It shows that the translation initiation may take place with some low efficiency in the absence of a conventional SD-interaction as observed also by others (reviewed by Sprengart and Porter, 1997).

Is there a selection for low affinity coat protein binding sites?

In wild type RNA the secondary structure of replicase TIR participates in the control of the replicase gene translation by serving as repressor - phage coat protein binding site. This same interaction was also thought to nucleate the phage encapsidation. Peabody (1997) demonstrated recently, however, that the coat protein - RNA operator interaction is not required for that purpose. Our data are in full agreement with this finding. Neither phage RNA mutants nor most of the adapted pseudorevertants (Fig.5) contain a replicase TIR structure needed for high affinity binding of coat protein.

Nevertheless, there is no reason to believe that the selection pressure in favor of an operator structure would be absent. The phage obviously gains some advantage from having a high affinity coat protein binding site in

replicase start region because genomic location and the strength of its interaction with coat protein are conserved among diverse RNA phages (Witherell *et al*, 1991) and our data support this view. During evolution of mutant A16 at least three different descendants derived and the winner, the revertant A16.4, possesses a replicase operator hairpin (Fig.5).

Though, unlike simplicity of the evolution of the SD-sequence the construction of a specific structure by chance mutations is much less likely because it requires several, possibly simultaneous, base changes. In addition the first rate requirement for ribosome binding with appropriate efficiency (SD-sequence) may provide another constraint on the evolution of a structure. Even so, wild type like hairpin reappeared in one instance (A16.4). What, however, turned out to be rather general pattern was creation of stem-loop structures that meet a requirement for the MS2 coat protein binding site. Such hairpins possess the loop of three nucleotides and the bulged adenosine in the stem or a tetraloop with the consensus sequence but lack the bulged adenosine at the required position. It is conceivable that coat protein when reach high concentration in the cell bind such hairpins in preference to many others RNA helixes.

What is the role of the coat protein-RNA operator interaction in light of its dispensability for the phage RNA encapsidation? Most probably the interaction provides a way to adapt the timing and level of the replicase synthesis to the phage needs. When the phage RNA replication has yielded amount of RNA sufficient for the protein synthesis and virion formation, the coat protein produced from these strands binds to the replicase gene start region. Thus the interaction may tell the phage that no more replicase is needed. Although the own phage RNA features may ensure an appropriate translational yield of the replicase protein they cannot ensure the coordination of its production to the expression of other phage genes as it is achieved by making the replicase synthesis dependent on the amount of the coat protein.

Is there selection against high level of the replicase synthesis?

Evolutionary studies carried out on the TIR of the MS2 coat protein gene showed that the phage RNAs which contained the mutations inhibiting the translational start upon evolution always reestablished its original efficiency. In the present study on the replicase gene TIR we did not found such phenomenon. The efficiency of the replicase gene translation in pseudorevertants varies around some 10% of the wild type phage. The question arises why this is so? There are at least three feasible explanations. One relates the low replicase initiation rate to some unknown evolutionary constraints implied by phage RNA features. For instance, we found that changes in sequence or structure of the replicase gene TIR may result in an exceeding suppression of the replicase translation by coat gene sequences. For this reason any phage attempt to increase the efficiency of the RBS itself may be contradicted by randomly arising decrease in the efficiency of the whole TIR due to appearance of some inhibitory RNA-RNA LDIs with the coat gene. More likely, however, are two other possibilities. First, a low replicase yield satisfies the phage RNA replication and there is no need to increase it further. Thus, also MS2 phage produces not much of replicase but in wild type situation this is achieved by coordinated replicase gene translation. Second, an efficient replicase gene translation is selected against by natural evolution. We envisage at least two reasons why it may be so. (i) Replicase is highly poisonous protein for the bacteria and its overproduction leads to cell death (Remaut *et al*., 1982; Berkhout and van Duin, 1985). (ii) Ribosomes and replicase compete for the common binding site on the phage RNA at the beginning of the coat cistron that is the only efficient ribosome entry site on intact phage RNA (Kolakofsky and Weissmann, 1971). A high continuos supply of replicase subunit would interfere

with the phage RNA translation.

If there is selection pressure against high replicase yield phage RNA mutants containing the replicase gene TIR that is transitionally as active as the wild type TIR but is not repressed by coat protein must evolve inhibition of replicase translation. Peabody (1997) had obtained such phages. They arisen to escape the super-repression with coat protein. To our knowledge their evolutionary adaptations upon prolonged growth in F^+ host have not been followed.

Tuning of the replicase gene TIR by long-term evolution

The coat protein-replicase gene TIR interaction has apparently been tuned by long-term natural evolution to achieve a sufficient strength but avoid super-repression (Peabody, 1997). Our data suggest that the replicase initiation site has been tuned also in another way, to avoid undesirably strong coupling to the coat gene translation. We have observed in this study and previously (Licis *et al.*, 1998) that changes in the RNA sequence and/or secondary structure of the replicase TIR somehow result in an increased suppression of replicase translation by the coat gene sequences. An essential aspect seems whether or not the replicase start AUG codon is located within a replicase initiator hairpin. We determined very strong replicase gene suppression for two revertants in which RNA the AUG codon was displaced outside initiator hairpin. In addition, spatial organization of the RNA might influence the level of suppression.

References

Barrick, D., Villanueba, K., Childs, J., Kalil, R., Schneider, T.D., Lawrence, C.E, Gold, L. and Stormo, G.D. (1994). Quantitative analysis of ribosome binding sites in *E.coli*. *Nucl. Acids Res.* **22**, 1287-1295.

Beckett, D. and Uhlenbeck, O. C. (1988). Ribonucleoprotein complexes of R17 coat protein and a translational operator analog. *J. Mol. Biol.* **204**, 927-938.

Berkhout, B. (1986). Translational control mechanisms in RNA bacteriophage MS2. Ph.D. Thesis, Leiden University.

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

Berkhout, B., de Smit, M.H., Spanjaard, A. Blom, T. and 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. *The EMBO J.* **4**, 3315-3320.

Bernardi, A. and Spahr, P.F. (1972). Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17. *Proc. Natl. Acad. Sci. USA,* **69**, 3033-3037.

Boni, I.V., Isaeva, D.M., Musychenko, M.L. and Tzareva, N.V. (1991). Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucl. Acids Res.* **19**, 155-162.

Calogero, R.A., Pon, C.L., Canonaco, M.A. and Gualerzi, C.O. (1988). Selection of the mRNA translation initiation region by *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. USA,* **85**, 6427-6431.

de Smit, M.H. and van Duin, J. (1990a). Control of prokaryotic translation initiation by mRNA secondary structure. *Progr. Nucl. Acid. Res. Mol. Biol.* **38**, 1-35.

de Smit, M.H. and 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.

de Smit, M.H. and van Duin, J. (1994). Translation initiation on structured messengers: Another role for the Shine-Dalgarno interaction. *J. Mol. Biol.* **235**, 173-184.

Dreyfus, M. (1988). What constitutes the signal for the initiation of protein synthesis on *Escherichia coli* mRNAs? *J. Mol. Biol.* **204**, 79-94.

Ganoza, M.C., Kofoid, E.C., Marliere, P. and

Louis, B.G. (1987). Potential secondary structure at translation-initiation sites. *Nucl. Acids Res.* **15**, 345-360.

Gold, L. (1988). Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Ann. Rev. Biochem.* **57**, 199-233.

Gold, L., Pribnow, D., Schneider, T.D., Shinedling, S., Singer, B.S. and Stormo, G.D. (1981). Translation initiation in prokaryotes. *Annu. Rev.Microbiol.* **35**, 365-403.

Hartz, D., McPheeters, D.S., Green, L. and Gold, L. (1990). Detection of *Escherichia coli* ribosome binding at translation initiation sites in the absence of tRNA. *J. Mol. Biol.* **218**, 99-105.

Hui, A. and de Boer, H.A. (1987). Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA,* **84**, 4762-4766.

Kolakofsky, D. and Weismann, C. (1971). Possible mechanism for transition of viral RNA from polysome to replication complex. *Nature New Biol.* **231**, 42-46.

Licis, N., Balklava, Z. and Berzins, V. (1998) Long-range translational coupling in single stranded RNA bacteriophages: an evolutionary analysis. *Nucl. Acids Res.* (Submitted for publication)

Lowary, P.T. and Uhlenbeck, O.C. (1987). An RNA mutation that increses the affinity of an RNA-protein interaction. *Nucl. Acids Res.* **15**, 10483-10493.

McCarthy, J.E.G. and Gualerzi, C. (1990). Translational control of prokaryotic gene expression. *Trends Genet.* **6**, 78-85.

Melancon, P., Leclerc, d., Destroismaisons, N. and Brakier-Gingras, L. (1990). The anti-Shine-Dalgarno region in *Escherichia coli* ribosomal RNA is not essential for the correct selection of translational starts. *Biochemistry,* **29**, 3402-3407.

Miller, J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 352-355.

Olsthoorn, R.C.L., Licis, N. and van Duin, J. (1994). Leeway and constraints in the forced evolution in a regulatory RNA helix. *The EMBO J.* **13**, 2660-2668.

Peabody, D.S. (1997) Role of the coat protein-mRNA interaction in the life cycle of bacteriophage MS2. *Mol. Gen. Genet.* **257**, 358-364.

Ravetch, J.V. and Jakes, K.S. (1976). Intact 3'-end of 16S rRNA is not required for specific mRNA binding. *Nature,* **262**, 150-153.

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

Remaut, E., de Waele, P., Marmenout, A. Stanssens, P. and Fiers, W. (1982) Functinal expression of individual plasmid-coded RNA bacteriophge MS2 genes. *The EMBO J.* **1**, 205-209.

Ringquist, S.T., Jones, T., Snyder, E.E., Gibson, T., Boni I. and Gold, L. (1995). High-affinity RNA ligands to *Escherichia coli* ribosomes and ribosomal protein S1: comparision of natural and unnatural binding sites. *Biochemistry,* **34**, 3640-3648.

Schmidt, B.F., Berkhout, B., Overbeek, G.P., van Strien, A. and van Duin, J. (1987). Determination of the RNA secondary structure that regulates lysis gene expression in bacteriophage MS2. *J. Mol. Biol.* **195**, 505-516.

Schneider, D., Tuerk, C. and Gold, L. (1992). Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. *Biochemistry,* **228**, 862-869.

Shine, J. and Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**, 1342-1346.

Sprengart, M.L. and Porter A.G. (1997). Functional importance of RNA interactions in selection of translation initiation codons. *Mol.Microbiol.* 19-28.

Stankevich, E., Kumpins, V., Licis, N., Klovins, J. and Berzins, V. (1995). Template activity of oligoribonucleotides synthesized by the phosphoramidite method using 2'-O-tetrahydrofuranyl and 2'-O-

tetrahydropyranyl protected groups. *Nucleosides and Nucleotides,* **14**, 1047-1048.

Tzareva N.V., Makhno, V.I and Boni, I.V. (1994). Ribosome-messenger recognition in the absence of the Shine-Dalgrno interactions. *FEBS Lett.* **337**, 189-194.

Uhlenbeck, O.C., Carey, J., Romaniuk, P.J., Lowary, P.T. and Becket, D. (1983). Interaction of R17 coat protein with its RNA binding site for translational repression. *J. Biomol. Str. Dyn.* **1**, 539-552.

van Himbergen, J., van Geffen, B. and van Duin, J. (1993). Translational control by a long-range RNA-RNA interaction; a basepair substitution analysis. *Nucl. Acids Res.* **212**, 1713-1717.

van Dieijen, G., van Knippenberg, P.H. and van Duin, J. (1976). The specific role of ribosomal protein S1 in the recognition of native phage RNA. *Eur. J. Biochem.* **64**, 511-518.

van Duin (1988). The single-stranded RNA bacteriophages. In *The Bacteriophages* (Calendar, R. ed.) vol.1, pp117-167, Plenum Press, New York.

Witherell, G.W., Gott, J.M. and Uhlenbeck, O.C. (1991). Specific interaction between RNA phage coat proteins and RNA. *Prog. Nucl. Acid Res. Mol. Biol.* **40**, 185-220.

Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science*, **244**, 48-52.

V

Long-range translational coupling in single stranded RNA bacteriophages: an evolutionary analysis

Normunds Licis, Zita Balklava and Valdis Berzins*
Biomedical Research and Study Centre, University of Latvia, Ratsupites 1, LV1067, Riga, Latvia
*Corresponding author

ABSTRACT

In the coliphage MS2 RNA a long-distance interaction (LDI) between the internal segment of the upstream coat gene and the start region of the replicase gene prevents the initiation of replicase synthesis in the absence of coat translation. Elongating ribosomes break up the repressor LDI and thus activate the hidden initiation site. Expression studies on partial MS2 cDNA clones led to identify the 1427-1433/1738-1744 LDI, the so-called Min Jou (MJ) interaction, as the molecular basis for the long-range coupling mechanism. Here, we verify the biological significance of this interaction for the control of the replicase gene translation in live phage. The LDI was disrupted by directed mutations in the 3'-side of the basepairing and the evolutionary adaptations of mutants were monitored upon phage passaging. Two categories of pseudorevertants derived. Those of the first type had restored the MJ-interaction but not necessarily the native sequence. The pseudorevertants of the second type evolved the compensatory substitutions downstream the MJ-interaction region and mostly contained a base change stabilising another 1419-1426/1810-1817 LDI located nearby the replicase initiation site. In one examined case we confirmed that the second-site mutations had restored the initially impaired coat-replicase coupling and repressed the replicase translation again.

INTRODUCTION

Translation initiation is usually the rate-limiting step in setting the level of gene expression in prokaryotes (1) and, as such, it represents the focal point of many regulatory mechanisms (2). Secondary structures in the messenger RNA can effectively restrict the accessibility of the initiation site (3). In the single stranded RNA bacteriophage MS2 the genomic RNA folding sequesters the start of the replicase cistron and the transit of ribosomes through the proximal part of the upstream coat gene is needed for its activation (4). It was proposed that in the unperturbed state of the viral RNA the replicase initiation site is hidden by a long range basepairing to the internal 1409-1433 coat gene segment (5). Ribosomes translating the coat message temporarily unfold the LDI that allows other ribosomes to bind at the replicase start. Firm support for the long-range model came from the finding that deletion of the 1419-1432 sequence abolished the coat-replicase coupling (6).

Recently, expression studies on partial MS2 cDNA clones (7) provided more distinct evidence in the role of the LDI in replicase repression. Disruption of the 1427-1433/1738-1744 MJ-interaction by substitutions on either side leads to increased replicase synthesis in the absence of coat translation whereas repair of the mismatches by compensatory base substitutions has the opposite effect. Here, taking an advantage of the ease with which RNA phage adapts to mutational perturbations, we analyse the MJ-interaction in live phage. Given its role in the translational control is important, phage mutants in which the basepairing is destroyed are expected to evolve suppressor mutations repressing the replicase translation.

MATERIALS AND METHODS

Bacterial strains and plasmids

Plasmid pMS2000 contains full-length MS2 cDNA under control of the thermoinducible P_L promoter of

phage λ (9), and bacteria transformed with pMS2000 produce viable phage. Vector pD plasmid carries linker TCTAGAgttaacggtaccCTTAAG (lower case letters refer to non-phage sequence) in place of the phage cDNA XbaI-BfrI (1303-1901) fragment. Cells transformed with this vector do not produce phage. Random PCR-directed mutagenesis was carried out on the phage XbaI-BamHI (1303-2057) cDNA fragment subcloned into the vector pTZ19 and the resulting plasmid was designated as pTN1. In plasmids used for the measurements of the replicase gene expression the MS2 cDNA fragments are located behind the P_L promoter, and *lacZ* gene is fused at the BamHI site to the replicase gene (7).

Clones containing phage λ P_L promoter were grown in *E.coli* K12 strain M5219 (M72 $trpA_{am}$, $lacZ_{am}$, $Sm^r/\lambda dbio_{252}$, $cI_{857}\Delta H1$) that encodes the thermosensitive λ repressor (cI_{857}) and the transcriptional antitermination factor N (10). *E.coli* AB259 F^+ cells (Hfr3000, Thi^-, Su^-) were used for amplification of phages. *E.coli* JM109 (*F'traD36 proAB lacI*$^q\Delta$*(lacZ) M15/endA1*) was the host strain for plasmid pTN1. Bacteria were grown on LB broth containing 10 g bactotryptone, 5 g yeast extract and 10 g NaCl per litre, and included an antibiotic if required.

Construction of vector plasmid pD

To construct pD plasmid 3 pmol of phosphorylated oligonucleotides CTAGAgttaacggtaccC and TTAAGggtaccgttaacT were combined, incubated at 70°C for 5 min and kept at room temperature for additional 15 min to produce a dsDNA linker with sticky XbaI and BfrI ends. The linker was then ligated to XbaI and BfrI digested plasmid pMS2000 DNA.

Mutagenesis and cloning

The oligonucleotides used in this study, donated by Dr. E. Stankevich, were synthesized on Gene Assembler and purified on C18 - Pep RPC HR 5/5 columns (Pharmacia). Complementary primers p52 and p53 were used to randomize the region of interest on MS2 cDNA by two-step PCR. Oligo p52 was synthesized as CGCCGGnnnnnnnAACATGAGGATTACCCATGT (parallel to the 1732-1764 sequence of MS2 genome) and oligo p53 as CGACATGGGTAATCCTCATGTTnnnnnnnCCGG (anti-parallel to the 1734-1766 sequence of MS2 genome), where 'n' stands for any nucleotide. In the first step two overlapping mutant MS2 cDNA fragments were created using primer pairs p52 - p109 (anti-parallel to the 2042-2062 sequence of MS2 genome) and p53 - reverse primer (standard primer for pTZ19). Reaction mixtures (100 µl) contained 5 µg/ml of pTN1 plasmid DNA, 0.5 µM of each primer pair, 200 µM dNTPs, 0.1 mg/ml BSA, 2.5 mM $MgCl_2$, 10 µl of 10x PCR buffer (100 mM TRIS-HCl, pH 8.8, 500 mM KCl, 0.8% NonidetP40) and 2 units of Taq DNA polymerase (Fermentas). The reaction was carried through 25 cycles: denaturation at 95°C for 45 sec, annealing at 40°C for 45 sec, and elongation at 72°C for 90 sec. The desired DNA fragments were then purified electrophoretically in a 6% polyacrylamide gel. In the second step, the two overlapping MS2 cDNA fragments from the first PCR were combined, and PCR with flanking p109 and reverse primer was carried out as above, except that the p52 and p53 primers as well as the pTN1 plasmid were omitted from the reaction mixture.

Mutants of the infectious clone were constructed by digesting the final PCR-product with XbaI and BfrI restriction endonucleases and then ligating them to the corresponding restriction sites in the vector plasmid pD. The ligated DNA was used to transform *E.coli* M5219 cells. Transformed bacteria were seeded onto nitrocellulose filters (HATF, pore size 45 µm (Millipore)) on top of LB agar plates containing kanamycin (50 µg/ml) and grown for 16-20 hours at 28°C. Plaque forming colonies were identified by replica plating the nitrocellulose filters on a lawn of *E.coli* AB259 F^+ cells on 0.8% agar and their cells were then inoculated into 3 ml of LB broth containing kanamycin (50 µg/ml), and grown overnight at 28°C. Dideoxynucleotide chain termination sequencing (11) of the target region on ds DNA was carried out with T7 DNA polymerase (Pharmacia) using primer p108 (parallel to the 1628-1648 sequence of MS2 genome).

Monitoring phage evolution and competition

Phages for the first infection were obtained after growth of the infectious phage cDNA clones overnight at 28°C. Diluted samples of the lysate containing about 10^5 plaque-forming units (pfu) of the phage were then inoculated in 3 ml of *E.coli* AB259 suspension (at OD_{670} of about 0.2), and the cultures were shaken overnight at 37°C. Diluted samples of the obtained lysate (about 10^5 pfu of the phage) were used for the next infection and so on. The infections were continued for 20-40 growth cycles. If the examination of the RNA sequence after various passages suggested genetic heterogeneity in the phage population, plating was used to separate individual genomes for subsequent sequence analysis.

To test the fitness of a mutant phage it was mixed with wild type phage in a pfu ratio of 1:1 and subsequently each mixture was allowed to complete two passages on F^+ bacteria. The phage RNA from the initial and the evolved samples were sequenced. The disappearance of the mutant RNA from the evolved sequence as compared to the initial one is defined as a failure of the mutant to compete with the wild type whereas the preservation of mutant RNA in the evolved sample is defined as good competition.

Sequence analysis of phage RNA

Phage RNA was prepared from 1 ml of lysate as

described by Olsthoorn *et al.* (9) and dissolved in 20 µl of water. 1-3 µl samples were taken for reverse transcription using primer p109 at 42°C for 30 min in 10 µl of 50 mM TRIS-HCl, pH 8.3, 4 mM KCl, 8 mM $MgCl_2$ and 10 mM DTT containing 1 unit of AMV reverse transcriptase (Pharmacia). The cDNA product from reverse transcription reaction was PCR-amplified with primers p109 and p180 (parallel to the 1200-1217 region of MS2 genome), and then selectively precipitated with one volume of ethanol in the presence of 2.5 M ammonium acetate (12). The amplified phage ds cDNA (bases 1200-2062 of MS2 cDNA) was directly sequenced with T7 DNA polymerase. Routinely, two oligonucleotides p108 and p105 (anti-parallel to the 1572-1585 region of MS2 genome) were used as primers for sequencing segments on either side of the MJ-interaction region which included bases 1700-1900 and 1300-1500 on MS2 cDNA, respectively. Other primers were also sometimes used to check the sequence of bases 1500-1700 in the phage genome.

Detection of replicase and lysis gene expression

E.coli M5219 cells that harbours the specified plasmids with MS2 cDNA sequences behind the P_L promoter were grown until A_{670} of about 0.25 was reached. The P_L promoter was then induced by shifting the cultures to 42°C. To measure the replicase expression cultures were put on ice after 30 minutes and the β-galactosidase activity was determined in 0.1 ml samples according to standard procedure (13). To monitor the lysis gene expression the cell density at A_{670} was recorded (14).

RESULTS

The experimental system

A complete cDNA copy of MS2 genome is present on a plasmid under transcriptional control of the thermoinducible promoter P_L of phage λ (9). The plasmid is maintained in *E.coli* M5219 F^- host cells, which produce phage spontaneously. Adverse mutations lower the titer of the phage. However, upon passaging mutants can restore high viability by base changes that arise due to a low copying fidelity of phage RNA replicases (15). These (pseudo)revertants may predominate through natural selection if compensation for the inflicted defect occurs.

The mutants

To disrupt the MJ-interaction in the phage RNA random or distinct nucleotide substitutions (mutants MJ1, MJ3-MJ22 and MJ2, respectively) were introduced in the infectious MS2 cDNA clone into the 3'-side of the LDI (1738-1744 sequence; Fig.1a and Tab.1). The 1738-1744 sequence codes for non-essential aminoacids of the lysis protein (16) and was therefore preferred over the coat segment of the long-distance basepairing as the target for base changes. The mutations lowered phage fitness as revealed by the decreased titers of the infectious MS2 cDNA clones (Tab.1) and confirmed by the failure of mutant MJ6 to survive in face of the wild type during co-infections.

Table 1. Properties of MJ-mutants.

Clone*	1738-1744 RNA sequence**	Titer,*** pfu/ml
Wild type	ccauuca	$1x10^{11}$
MJ6	ccUAGAC (5)	$4x10^{10}$
MJ18	ccUAuGC (4)	$4x10^{10}$
MJ14	ccaGCcC (3)	$4x10^{10}$
MJ8	AUUCucC (5)	$3x10^{10}$
MJ10	cUUCuGC (5)	$3x10^{10}$
MJ3	GcCCAcG (5)	$3x10^{10}$
MJ20	ccGACGG (5)	$3x10^{10}$
MJ16	ccUGuGC (4)	$3x10^{10}$
MJ9	cUCGCUG (6)	$3x10^{10}$
MJ2	ccUuuGa (2)	$2x10^{10}$
MJ15	GUUuAcC (5)	$2x10^{10}$
MJ21	cACuuGC (4)	$2x10^{10}$
MJ4	UcCuucC (3)	$2x10^{10}$
MJ1	cAUGuUa (4)	$2x10^{10}$
MJ7	UUCGAcC (6)	$1x10^{10}$
MJ17	AcGACUG (6)	$1x10^{10}$
MJ12	cUGCAUa (5)	$7x10^{9}$
MJ22	cGGCCcC (5)	$7x10^{9}$
MJ19	UUUGGAa (6)	$4x10^{9}$
MJ11	cUGuCcC (4)	$2x10^{9}$
MJ5	UcGuCca (3)	$2x10^{9}$

*Wild type construct is pMS2000 (9). The mutants were created by PCR directed mutagenesis using a primer with a randomized sequence corresponding to the 1738-1744 phage segment, except for MJ2 which was derived from the corresponding partial MS2 cDNA clone (7).

**The 1738-1744 RNA sequence of the mutants is aligned with that of the wild type. Capital letters indicate the mutations. Numbers in parentheses show the number of base changes as compared to native RNA.

***Appropriate dilutions of supernatant from cDNA cultures grown overnight at 28°C were tested on lawns of *E.coli* AB259 F^+ cells.

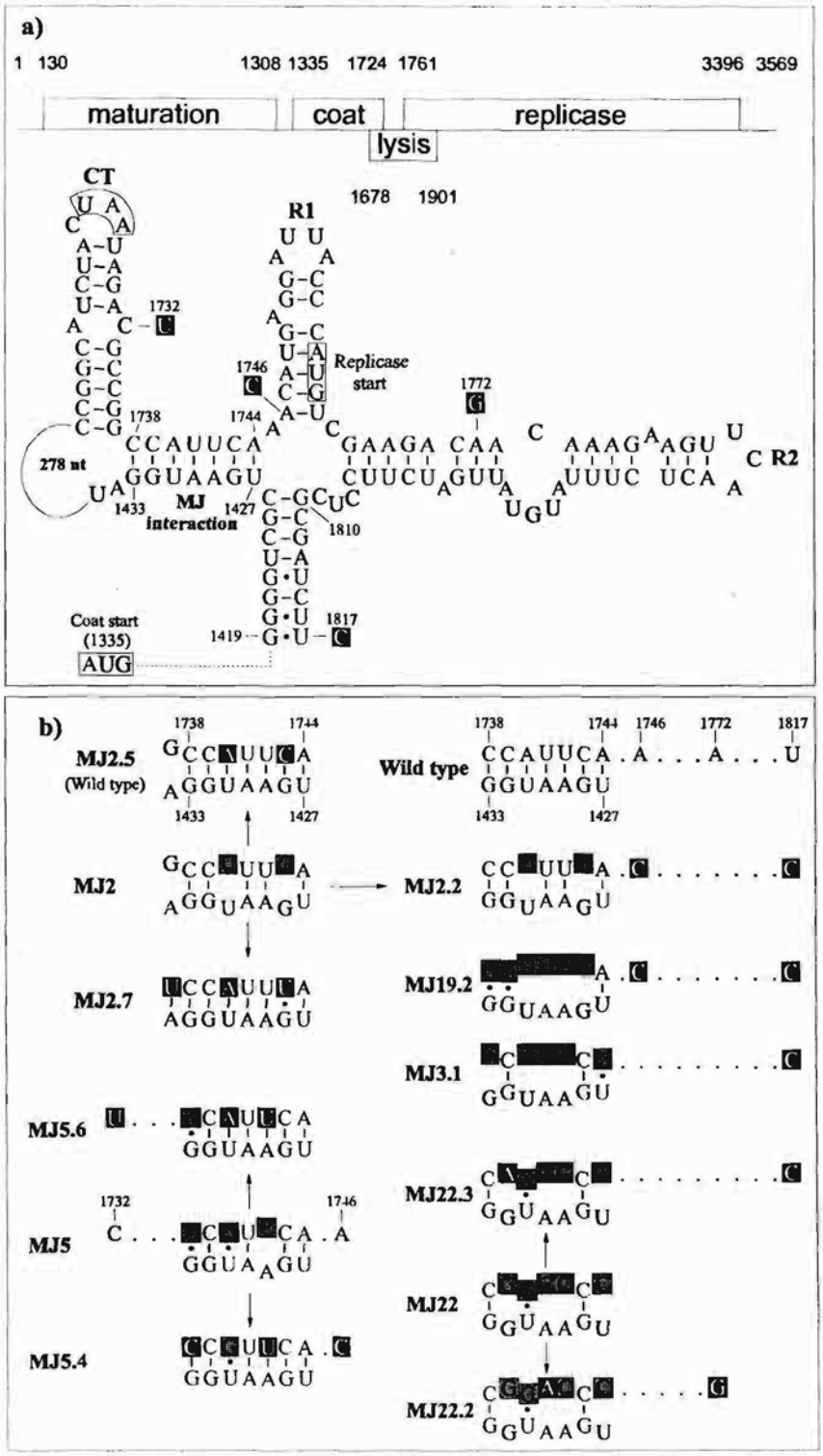


Figure 1. (a) Genetic map of RNA coliphage MS2 and RNA secondary structure model for the replicase initiation site (8). The start codon of the replicase and the stop codon of the coat gene is marked by a box. Local stem-loop structures are indicated as CT, R1 and R2. (b) Comparison of wild type, mutant and pseudorevertant phage RNA sequences and structures of MJ-interaction region. The designation of various phages depicts their origin. For instance, MJ2.2, MJ2.5 and MJ2.7 derived from mutant MJ2. The initial mutations are in grey while suppressor mutations in black boxes. Left column shows the reconstruction of MJ-interaction. The right column compiles the descendants, which did not evolve the original LDI. Here, the MJ-interaction region is shown for illustrative purposes only. The base changes outside the MJ-interaction region are indicated. They are also shown on secondary structure model of MS2 RNA in Fig.1a. The MJ-interaction region in phages MJ3.1 and MJ19.2 is the same as in their parental mutants which are therefore omitted.

From about twenty mutants collected five (MJ2, MJ5, MJ3, MJ19, MJ22) were taken for evolutionary analysis. One criterion that underlined the choice of the above mutants was the number of the nucleotide substitutions. MJ2 was the single mutant containing only two base changes and MJ5 also possessed only a few (three) substitutions. We considered that they provide the phage with a reasonable chance to re-establish the original regulatory LDI. Conversely, the many substitutions in MJ3, MJ22 and MJ19 RNA (five and six changes, respectively) may render restoration of the wild type pairing highly unlikely (14) thereby forcing the phage to find an alternative to keep the replicase expression under control. On the other hand, mutants MJ5, MJ19 and MJ22 were preferred over others because they showed a low initial titer that usually speeds up evolution whereas MJ3 was a random choice. The course of the evolutionary adaptation was followed for many growth cycles as described in Materials and Methods.

Reconstruction of MJ-interaction

Several descendants of mutants MJ2 and MJ5 evolved back the MJ-interaction (Fig.1b, left column) and one outcome of such adaptation was wild type phage (revertant MJ2.5) derived from MJ2 in a stepwise manner. This pathway obviates the advantage of wild type but is inconclusive otherwise. An informative revertant of the MJ-interaction is MJ2.7, which shows that the secondary structure of the phage RNA but not the sequence is important. Here, three base changes re-established the LDI while only one of them is a true reversion and the G1737U substitution resulted in the extra U-A basepair of the interaction. The significance of the LDI is also implied by the fact that all of the suppressor mutations in MJ2.7 as well as MJ2.5 are transversions although the replication errors are predominantly transitions (14).

Two revertants (MJ5.4 and MJ5.6) arose from mutant MJ5. MJ5.6 evolved a nearly native LDI but still preserved the U.G pair where in wild type is C-G. On the other hand, the C1732U second-site substitution in the phage turned the mismatch of the coat terminator (CT) hairpin into an A-U basepair. The CT hairpin possibly stabilises the MJ-interaction by coaxial stacking (7) and thus strengthening of the helix may help to suppress the replicase translation. Perhaps, closing the mismatch is sufficient to compensate for a small defect in the interaction. Less straightforward is revertant MJ5.4. This phage besides two reversions to the LDI also acquired an unclear A1746C substitution in the replicase initiator R1 hairpin.

Progeny phages without a MJ-interaction

The progeny of MJ3, MJ19 and MJ22 as well as one descendant of MJ2 did not restore the MJ-interaction (right column at Fig.1b) and their evolutionary adaptations strikingly resembled by the U1817C base change some 50 nucleotides downstream the replicase start (MJ3.1, MJ2.2, MJ19.2 and MJ22.3). Pseudorevertants MJ2.2 and MJ19.2 contained also in common the A1746C substitution. Probably, these second site mutations ensure a good compensation for the absence of the MJ-interaction but alternatives may also exist as illustrated by MJ22.2.

To assess whether the MJ-unrelated substitutions interfered with the replicase translation the 1365-2057 cDNA fragment of MJ19.2 and also the corresponding region of MJ19 were fused to the *lacZ* gene in an expression vector (7). Data on these clones (Tab.2) showed that the mutations evolved in MJ19.2 decreased about six times the replicase synthesis and thus fully compensated for the inflicted activation of the replicase initiation site in mutant MJ19. In fact, MJ19.2 produce even less replicase than does wild

Table 2. Relative expression of the replicase gene.

RlacZ	MS2 cDNA fragment		Coupling*
	1365-2057	1628-2057	
Wild type	1.0	9.0**	9
MJ19	3.1	9.2	3
MJ19.2	0.5	4.0	8

The efficiency of replicase gene translation is measured as the activity of replicase-β-galactosidase fusion protein (RlacZ) using clones that contain either 1365-2057 or 1628-2057 MS2 cDNA fragment fused to the lacZ gene at BamHI site. The constructs lack the start of phage coat cistron. In plasmids, which begin at position 1365 of MS2 cDNA, the coat gene segment that couples the translation of replicase in wild type phage is present (6). The RlacZ activity of wild type clone (7) is set at 1.0 (note that coupling is not absolute (4)). In constructs starting at position 1628 most of the coat gene is absent and replicase synthesis is uncoupled. Values are averaged from 5 experiments.

*Coupling is calculated as the ratio between the RlacZ activities of the corresponding clones starting at 1628 and 1365.

**Data of Berkhout and van Duin (6)

type. These results demonstrate that evolution of pseudorevertants was directed towards suppression of the replicase translation. The data, however, leave out the question if the evolved mutations affected the coat-replicase coupling. To resolve the point the 1365-1628 sequence was deleted from the experimental constructs and the replicase synthesis in the resulting clones was measured (Tab.2; clones starting at 1628). Because the removed region involves the essential part of the coat gene the replicase expression in the shortened variants is no longer under the coat control and thereby goes up as compared to the corresponding parental clones. This increase gives a quantitative measure for the coupling (coupling value, Tab.2). In wild type the deletion of the coat cistron rise replicase synthesis about nine times (6). In MJ19 we expect an impaired long-range control due to the disruption of the MJ-interaction and this expectation is burn out by a low coupling value. The most important finding is that in MJ19.2 the suppressor mutations had restored the dependence of the replicase gene translation on that of the coat. Thus, although the phage did not revert to the MJ-interaction it did revert to the coupling phenotype.

The question arises how the replicase translation is regulated in MJ19.2? Most probably some LDI(s) involving a coat gene segment mediate the control. The U1817C exchange strengthens the 1419-1426/1810-1817 LDI (Fig.1a). The interaction is located nearby the replicase initiation site and may interfere with the ribosome binding. As discussed further this is also suggested by the data on the replicase expression in partial cDNA clones.

The A1746C substitution and also the evolved mutations in MJ22.2 might set off different basepairing(s) between the replicase initiation site and the coat gene as revealed by computer analysis (17) of the phage RNA structures (data not presented). Yet, we cannot reliably account for such models. Also, the presence of the A1746C substitution in revertant MJ5.4 seems to contradict a 'LDI explanation'. At present, this and some other mutations are poorly understood.

Other points

To verify that MJ19.2 did not contain second site substitutions outside the monitored sequences the 1303-1901 fragment was cloned back in the full-length MS2 cDNA copy. The titer of the infectious clone has increased about 25 times as compared to the parental MJ19 clone. Also, the phage produced from the plasmid competed well with the wild type during co-infections. These results show that the two suppressor mutations renewing the translational control are also entirely responsible for a high viability of MJ19.2. To exclude the effects of the mutations on the lysis function the L gene expression was monitored both in the full-length and the partial MJ19 and MJ19.2 cDNA clones and no differences from the wild type were found.

DISCUSSION

The synthesis of replicase in the RNA coliphage MS2 is kept under dual control. One is repression by phage coat protein binding to the replicase initiator hairpin (18). The other is translational coupling to the upstream coat cistron that was proposed (7) to be effectuated by the long distance MJ-interaction (Fig.1a). Here, we provide evolutionary evidence in the biological significance of this LDI for the regulation of the replicase gene expression. Phage mutants in which the MJ-interaction was disrupted by mutations and thereby the replicase initiation site was activated evolved the suppressor mutations inactivating the translational start again. In four of the derived progeny phages it was accomplished by (pseudo)reversions to the original regulatory MJ-interaction. Five other obtained descendants did not evolve the native structure but acquired mostly the second site substitutions downstream the 3'-side of the MJ-interaction region. We assessed the replicase synthesis in one typical phage of the last type and found that the control of the replicase translation gene had been restored by the second site mutations. This result both supports the role of the MJ-interaction in the replicase gene regulation and shows the importance of the coupling mechanism. The coupling most probably serve to prevent uncontrolled access of ribosomes to the phage RNA that is needed to ensure the switch between phage RNA translation and replication (19).

Molecular basis for the coupling mechanism

The disruptions of the MJ-interaction by mutations to a large extent impair the dependence of replicase

translation on that of the coat gene (7). Still, they do not lead to fully uncoupled replicase synthesis as compared to the removal of the internal 1419-1432 coat cistron region or larger deletions (6,20; see also Tab.2). These observations imply the participation of the other phage RNA structure(s) in the control mechanism.

In an improved model of MS2 RNA secondary structure (8) the replicase initiation site is enclosed by two long distance basepairings to the 1419-1432 coat segment: the 1427-1433/1738-1744 MJ-interaction from one side and the 1419-1426/1810-1817 LDI from the other. It is reasonable that the last interaction also hinder the replicase initiation. Two kinds of our data substantiate this view. First, in the present study for out of nine (pseudo)revertants evolved the U1817C substitution (Fig.1a) that stabilises the 1419-1426/1810-1817 LDI. Second, two mutations (G1421A and C1424A) in the coat gene that disrupt the LDI activate about twofold the replicase expression in a partial MS2 cDNA clone in the absence of coat translation. Also, some pseudorevertants derived from the phage containing the two mutations evolved the U1817C base change in the replicase gene (our unpublished data).

The adjacent RNA structures might affect the stability of the LDIs. In particular, it was proposed that the coat terminator helix stacks upon MJ-interaction and increases its stability and thereby may influence the coupled replicase expression (7). In agreement with the above a partial revertant of the MJ-interaction acquired a second site mutation that strengthen the CT hairpin.

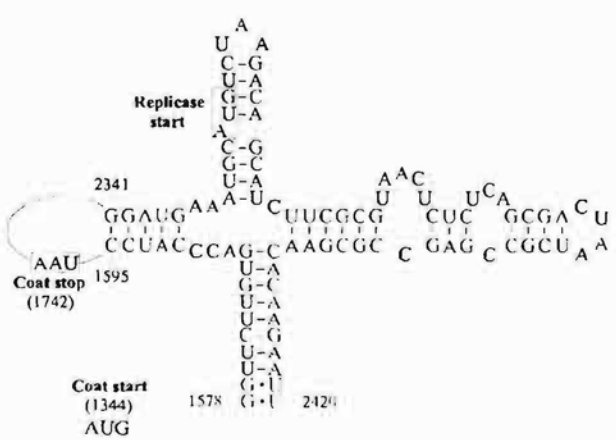


Figure 2. Secondary structure model for the translation initiation site of the replicase gene of RNA coliphage Qβ

Phylogenetic comparison

A comparable translational regulation of the replicase gene operates in distantly related phage Qβ (21). The RNA secondary structure proposed for the replicase initiation site (22,23; Fig.2) closely resembles the MS2 model. The structure adopts the same four-way junction framework containing two local hairpins and suggestive features of the model are the two LDIs to the upstream coat cistron. In Qβ these LDIs involve a segment near the end of the coat gene and may explain coat-replicase translational coupling (22).

ACKNOWLEDGMENTS

We thank Dr. Jan van Duin for his critical and helpful discussions. This work was supported by Latvian Science Council grant Nr. 93.563 to V.B.

REFERENCES

1 Gualerzi,C.O. and Pon,C.L. (1990) *Biochemistry*, **29**, 5881-5889.

2 Gold,L. (1988) *Annu. Rev. Biochem.*, **57**, 199-233.

3 de Smit,M.H. and van Duin,J. (1990) *Progr. Nucl. Acid Res. Mol. Biol.*, **38**, 1-35.

4 Hirouchi,K. (1975) In Zinder,N.D. (ed.), RNA Phages, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 29-50.

5 Min Jou,W., Haegeman,G., Ysebaert,M. and Fiers,W. (1972) *Nature (Lond.)*, **237**, 82-88.

6 Berkhout,B. and van Duin,J. (1985) *Nucleic Acids Res.*, **13**, 6955-6967.

7 van Himbergen,J., van Geffen,B. and van Duin,J. (1993) *Nucleic Acids Res.*, **21**, 1713-1717.

8 Groeneveld,H. (1997) Ph.D. thesis, Leiden University, Leiden, The Netherlands.

9 Olsthoorn,R.C.L., Licis,N. and van Duin,J. (1994) *The EMBO J.*, **13**, 2660-2668.

10 Remaut,E., de Waele,P., Marmenout,A. Stanssens,P. and Fiers,W. (1982) *The EMBO J.*, **1**, 205-209.

11 Sanger,F., Nicklen,S. and Coulson,A. (1977) *Proc. Nat. Acad. Sci. USA*, **74**, 5463-5467.

12 Landweber,L.F. and Kreitman,M. (1993) In Wu,R. (ed.) *Methods in Enzymology*, Academic Press, San Diego, Vol. 218, pp. 17-26.

13 Miller,J.H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

14 Klovins,J., Tsareva,N.A., de Smit,M.H., Berzins,V. and van Duin,J. (1997) *J. Mol. Biol.*, **265**, 372-384.

15 Domingo,E., Martinez-Salas,E., Sobrino,F., de la Torre,J.C., Portela,A., Ortin,J., Lopez-Galindez,C., Perez-Brena,P., Villanueva,N., Najera,R., vandePol,S., Steinhauer,D. de Polo,N., and Holland,J.J. (1985) *Gene*, **40**, 1-8.

16 Berkhout,B., de Smit,M.H., Spanjaard,R.A., Blom,T. and van Duin,J. (1985) *Nucleic Acids Res.*, **4**, 3315-3320.

17 Zuker,M. (1989) *Science*, **244**, 48-52.

18 Witherell,G.W., Gott,J.M. and Uhlenbeck,O.C. (1991) *Progr. Nucl. Acid Res. Mol. Biol.* **40**, 185-220.

19 Kolakofsky,D. and Weissmann,C. (1971) *Nature New Biol.*, **231**, 42-46.

20 Berkhout,B. (1986) Ph.D. thesis, Leiden University, Leiden, The Netherlands.

21 Ball,L.A. and Kaesberg,P. (1973) *J. Mol. Biol.*, **74**, 547-562.

22 Skripkin,E.A. and Jacobson,A.B. (1993) *J. Mol. Biol.*, **233**, 245-260.

23 Beekwilder,J. (1996) Ph.D. thesis, Leiden University, Leiden, The Netherlands.