Activation of RegB endoribonuclease by S1 ribosomal protein requires an 11 nt conserved sequence

Sylvain Durand, Graziella Richard, Marco Bisaglia[1], Soumaya Laalami, François Bontems[1] and Marc Uzan*

Institut Jacques Monod. UMR7592 CNRS-Universités Paris 6 and Paris 7. 2, Place Jussieu, 75251 Paris Cedex 05 and [1]ICSN-RMN, Institut de Chimie des Substances Naturelles, CNRS, 91190 Gif sur Yvette and Ecole Polytechnique, 91128 Palaiseau, France

ABSTRACT

The T4 RegB endoribonuclease cleaves specifically in the middle of the -GGAG- sequence, leading to inactivation and degradation of early phage mRNAs. *In vitro*, RegB activity is very weak but can be enhanced 10- to 100-fold by the *Escherichia coli* ribosomal protein S1. Not all RNAs carrying the GGAG motif are cleaved by RegB, suggesting that additional information is required to obtain a complete RegB target site. In this work, we find that in the presence of S1, the RegB target site is an 11 nt long single-stranded RNA carrying the 100% conserved GGA triplet at the 5′ end and a degenerate, A-rich, consensus sequence immediately downstream. Our data support the notion that RegB alone recognizes only the trinucleotide GGA, which it cleaves very inefficiently, and that stimulation of RegB activity by S1 depends on the nucleotide immediately 3′ to -GGA-.

INTRODUCTION

Messenger RNA degradation is a regulated process that contributes greatly to establishing the pattern of gene expression. Endoribonucleases play a crucial role in this process. In prokaryotes, endoribonucleolytic attacks trigger mRNA decay by creating entry sites for very active 3′–5′ exonucleases (1–3). In mammals, the list of endonucleases triggering the degradation of specific transcripts is growing rapidly (4). In spite of their central role in maturation and degradation of mRNAs, only a few endoribonucleases have been identified and described in detail so far. One of them, RNase E of *Escherichia coli*, plays a general role as it affects most cellular transcripts. Others, like RNase III, play more specific roles (5,6).

The sequence-specific, RegB endoribonuclease, encoded by bacteriophage T4, falls into the latter group (7,8). This RNase cuts T4 early mRNAs in the middle of GGAG and, very rarely, GGAU sequences, located in intergenic regions. The RNA must be single-stranded. Several lines of evidence suggest that its catalytic mechanism is like that of RNase T1 (9,10). The *regB* gene is widely spread among T4-related phages and shows strong sequence conservation (11,12).

During T4 phage infection, there are two consequences of RegB cleavages: (i) Functional inactivation of many early transcripts. This is due to the fact that the GGAG motif is one of the most frequent Shine–Dalgarno sequences encountered in T4 (13). (ii) Degradation of most early but not middle or late mRNAs (14). Thus, the RegB nuclease down-regulates the translation of many pre-replicative T4 genes. However, synthesis of a few middle proteins is stimulated by RegB (14). By providing a mechanism that frees the translation apparatus from abundant early mRNAs, RegB facilitates the transition between early and subsequent phases of T4 gene expression (8,14).

The RegB endonuclease requires a co-factor to act efficiently. When assayed *in vitro*, RegB activity is extremely low but can be stimulated up to 100-fold, depending on the RNA substrate, by the ribosomal protein S1 (15,16). The *E.coli* ribosomal protein S1 is an RNA-binding protein that plays an essential role during translation, clamping the ribosomes to mRNAs (17,18). The S1 protein contains six homologous RNA-binding domains, each of about 70 amino acid, called S1 domains (or S1 modules). The two N-terminal domains are involved in binding to the ribosome while the four C-terminal domains are devoted to mRNA interactions. Similar domains can be found in a large number of RNA-associated proteins from bacteria to humans, in particular in certain RNases (*E.coli* RNase E, polynucleotide phosphorylase and RNase II) and RNA-binding proteins of yeast and archaeal exosomes (19–22). The role of the S1 ribosomal protein is not limited to translation. S1 seems to be a multifunctional protein involved in several unrelated processes that all

*To whom correspondence should be addressed. Tel: 33-1 44277973; Fax: 33-1 44275716; Email: uzan@ijm.jussieu.fr
Present address:
Marco Bisaglia, Institute for Biomolecular Chemistry, CNR, Department of Chemical Sciences, University of Padova, Via F. Marzolo 1, 35131, Padova, Italy

use its RNA-binding and, possibly, single-stranded DNA-binding properties (8,17). The RegB and S1 protein do not seem to interact directly (8). In addition, RegB has a very low affinity for its RNA substrates [see (10,23); S. Durand, G. Richard and M. Uzan, unpublished data]. Therefore, it is highly probable that at least the first step in the S1 activation pathway of RegB involves S1 interactions with RNA.

As RegB is easily inhibited by secondary structures [(7,16); this work], one possibility would be that S1 stimulates RegB through its RNA unwinding ability (24,25). However, Lebars *et al.* (16) presented evidence that this is not the case. Using small, artificial RNAs, Bisaglia *et al.* (26) found that the entire S1 protein is not necessary to achieve efficient stimulation of the RegB reaction. The portion of the protein made up of the four C-terminal modules 3-4-5-6 mimics perfectly the effect of the whole protein. Depending on the substrate, domain 6 could be removed without affecting the extent of stimulation. The smallest domain combination able to stimulate the cleavage reaction significantly was the bi-module 4-5. Quite similar results were obtained with natural T4 mRNAs (S. Durand and M. Uzan, unpublished data). This suggests that domains 3, 4 and 5 contribute, directly or indirectly, to forming an RNA-binding surface able to interact with nucleotide of the RegB target sites.

During infection by bacteriophage T4, the action of RegB is limited to GGAG motifs located in intergenic regions of early transcripts. The GGAG motifs located in coding sequences or in intergenic regions of late mRNAs are not (or very poorly) cut. Therefore, the GGAG/U sequence is not sufficient for RegB recognition. *In vitro* as well as in uninfected *E.coli* cells where RegB is provided from a plasmid and the RNAs to be tested (T4 sequences) from another plasmid, the pattern of cleavage is identical to that found during T4 infection. This shows that the information determining whether RNA is cleaved or not by RegB is carried by the RNA, in the vicinity of the RegB target site (14). A SELEX experiment based on the selection of RNA cleaved by RegB in the presence of the S1 protein was carried out to identify the missing sequence or structural element (23). The selected sequences share the following properties. The GGAG tetranucleotide are common to all of them (one exception was GGAC), with a strong bias towards a GGAGG motif. In most cases, the tetra or pentanucleotide is located at the 5′ end of the randomized region, suggesting that the region 3′ to the GGAG(G) motif plays a role. This region is enriched in A and C nucleotide. However, the absence of any other clearly conserved sequence or structural motif besides -GGAG- prevented the authors from finding the missing piece of information that would have completed the description of a typical RegB site.

Lebars *et al.* (16) found that a small (30 nt long), structured RNA derived from one of the RNAs isolated in the SELEX experiment of Jayasena *et al.* (23) is efficiently and specifically cleaved by RegB in the absence of S1. Analysis by NMR spectroscopy indicated that the 3′ G of the GGAG motif is base paired while the rest of the motif is unpaired, in a loop. Any change in the sequence that perturbed this structural arrangement resulted in decreased cleavage efficiency. This suggests that, when S1 is required for efficient cleavage (with the natural T4 transcripts, for instance), the role of this protein is to promote this constraint on the

RNA. However, simple application of Watson–Crick base pairing rules does not allow one to fold the well-cleaved natural RNA molecules into the sort of structure exhibited by the small RNA analyzed by Lebars *et al.* (16). Thus, the question of RegB specificity still remained unanswered. In the present work, we show that the RegB target site spans a region of 11 nt of single-stranded RNA carrying the 100% conserved GGA triplet at the 5′ end and a degenerate consensus sequence immediately downstream, required for RegB stimulation by the S1 protein.

MATERIALS AND METHODS

Bacteriophages, bacteria and T4 infection conditions

Bacteriophage T4 wild-type was T4D (T4$^+$). The T4*regB-mis52* phage contains the *regB*K52L missense mutation (14). T4K10 (*38*amB262, *51*amS29 *denA*nd28 *denB-rIIB* ΔrIIPT8) was the recipient phage for *in vivo* directed insertion/substitution mutagenesis (27). Phage λCE6 (which carries the T7 RNA polymerase gene) (28) was amplified in ED8689 (*sup0*) and titered on Y-MC (*supF*) (29). *E.coli* B^E was the host for T4 infections. *E.coli* K12 CR63 (*supD*) was used to grow nonsense T4 phages. *E.coli* cells to be infected by T4 were grown at 30°C in MOPS-Tricine medium (30) supplemented with 0.4% (w/v) glucose and 1% casaminoacids. Infections were carried out at a multiplicity of 7, at a cell density of 5×10^8/ml.

Oligonucleotides

More than 70 oligonucleotides were used in the primer extension experiments designed to detect RegB cleavages in T4 RNAs. Their list and their sequences are available on request. Here, we present only the sequence of the oligonucleotides used in primer extension experiments shown in detail in this article.

denV: CACGAACACGTTTACCGTTAGCAAC, complementary to 103–128 nt of the *den*V coding sequence, mobD.5: CCATTGTCGTCTAATTCTAGC, complementary to 63–84 nt of the mobD.5 coding sequence.

Oligodeoxyribonucleotide labeling, primer extension and RNA sequencing

5′ End-labeling of oligodeoxynucleotides and RNA sequencing were performed as described in (9,13). Primer extension experiments were carried out essentially as described by Uzan *et al.* (9) except that Moloney murine leukemia virus reverse transcriptase (Invitrogen) was used (50 U per assay) for 50 min at 42°C, in the presence of RNasin RNase inhibitor (Promega).

Protein purification

His-tagged RegB nuclease was purified from the overproducing *E.coli* strain, JM101 (pARNU2), as described in Sanson *et al.* (14). The S1 ribosomal protein, His-tagged on its C-terminus, was purified under denaturing conditions from the overproducing strain BL21(λDE3) (pLysS; pET-rpsA), essentially as described by Wower *et al.* (31). The procedure includes two successive affinity chromatography columns: on Ni-agarose and then poly(U)-Sepharose. At the end of the

procedure, the S1 protein was dialyzed against 10 mM Tris–Cl (pH 7.4), 50 mM NH_4Cl, 0.1 mM Na_2EDTA, 7 mM β-mercaptoethanol and kept at 4°C. We found that the His-tagged S1 protein has the same stimulatory effect on RegB as the native form isolated from the ribosomes. T7 RNA polymerase was purified from the overproducing strain *E.coli* BL21 (pAR1219-T7 gene 1) according to the protocol of Zawadzki *et al*. (32).

RegB *in vitro* assay

Unless otherwise specified, the RegB cleavage reaction mix contained 1 μM ^{32}P-labeled RNA, 0.1 μM RegB, 0.3 μM S1 in 50 mM Tris–HCl (pH 7.0), 1 mM DTT and 0.1 mM Na_2EDTA. Incubation was at 37°C for the time indicated in each experiment. The reaction was stopped by the addition of an equal volume of 8 M urea, 1× Tris–borate–EDTA, 25 mM EDTA, 0.05% xylene cyanol and incubation at 90°C for 5 min. Samples were run on polyacrylamide (12%)–urea sequencing gels. The gels were analyzed by a Molecular Dynamics Storm 860 Phosphorimager. The percentage of cleavage was determined by calculating the ratio of radioactivity in the 'processed' band to the total amount of radioactivity in both the full-length RNA band and the processed species. For this purpose, ImageJ software was used (33).

In vitro transcription, RNA purification and labeling

The sequences of the templates used in this study are available on request. Annealing between the single-stranded template and the oligonucleotide complementary to the T7 promoter was done in TES [10 mM Tris (pH 7.5), 1 mM EDTA and 150 mM NaCl]. After heating at 95°C for 4 min, the mixture was slowly cooled down to room temperature. The transcription reaction mix contained 5 mM DTT, 40 mM Tris (pH 7.5), 1 mM spermidine, 0.01% Triton X-100, 20 mg/ml polyethylene glycol 8000, 20 mM $MgCl_2$, 5 mM NTP, 100 U of RNasin, 3 μM T7 RNA polymerase and 2.5 μM annealed template. The mix was incubated 4 h at 37°C. RNA was purified with phenol/chloroform pH 5, ethanol precipitated and loaded on polyacrylamide (12%)–urea gels. RNA was electro-eluted from the gel with the BIOTRAP system (Schleicher and Schuell) as indicated by the supplier. The RNAs were dephosphorylated with Calf Intestine alkaline phosphatase (Roche) and 5′ labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. When required RNA (100 pmol) was labeled in 3′ with ^{32}P-pCp (3.3 pmol, 10 μCi; Amersham) using 20 U of T4 RNA ligase (BioLabs) in 20 μl for 1 h at 37°C. The buffer was that provided by the supplier [50 mM Tris–HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP] to which 10% dimethyl sulfoxide (DMSO) and 20 U RNasin were added.

Enzymatic probing of RNA structure

RNA structures were probed with RNases T1 (Fermentas), T2 (Sigma), from 3.6×10^{-4} to 7.2×10^{-4} U/μl, A (Fermentas) from 2×10^{-7} to 4×10^{-7} μg/μl, and V1 (Pharmacia Biotech) from 3.6×10^{-4} to 7.2×10^{-4} U/μl. The reactions were carried out at 25°C for 20 min in the RegB reaction buffer. $MgCl_2$ (1 mM) was added with RNase V1. The pattern of cleavage by the other RNases was not influenced by the addition of 1 mM $MgCl_2$. RNA ladders were obtained by

incubating 80 pmol of RNA labeled in 5′ or 3′ with 60 mM Na_2CO_3, 0.4 mM EDTA and 3 μg of yeast tRNA in 25 μl final volume for 7 min at 80°C.

In vivo mutagenesis

PCR DNA fragments, 120 bp long, carrying one of the three possible mutations in the fourth position of the GGAG motif of *motB*, were cloned into plasmid pBR322. *E.coli* B^E cells transformed with each of the resulting plasmids were then infected by T4$^+$ in order to transfer the mutation to the T4 genome by homologous recombination. The insertion of the A-rich sequence of *motB* into the *denV* sequence to obtain the *denV*[AU] mutant (see Results) was performed using the directed Insertion/Substitution mutagenesis method described by (27). The T4 K10 *denV*[AU] recombinant was back-crossed three times with T4 wild-type phage. Recombinant phages were detected by plaque hybridization with a ^{32}P-labeled oligonucleotide complementary to the mutated sequence, as described in Sambrook *et al*. (34).

RESULTS

Additional sequence 3′ but not 5′ of the GGAG motif is required for efficient cleavage by RegB in the presence of S1

As shown by Sanson *et al*. (14), additional information determining whether a GGAG motif is cleaved or not is contained in *cis* in the vicinity of the cleavage site. In a first approach to identifying such elements, we determined the minimum length of RNA molecule able to be cut by RegB in the presence of the S1 protein.

After T4 infection, *motB* mRNA is efficiently cleaved by RegB in a GGAG motif (13). A 49 nt long truncated version of the *motB* transcript (Figure 1), hereafter called *motB* RNA, obtained by *in vitro* transcription, is also efficiently cut by RegB in the presence of S1 (see Figure 4A). The RNA products are 18 (5′ part) and 31 nt long. *motB* RNA was labeled in either 5′ or 3′ and then treated with sodium bicarbonate under mild conditions so as to generate all possible RNA sizes. The RNA mixture was then submitted to RegB/S1 digestion and samples of the reaction were withdrawn at time intervals for analysis by gel electrophoresis. Figure 2A shows the result of an experiment carried out with 5′ end labeled *motB* RNA. Cleavage by RegB results in the accumulation, over time, of the 18 nt RNA species. It can be seen that the longer the molecules, the faster they are cleaved by RegB/S1. However, after 40 min of reaction, there is essentially no further evolution of the pattern; a set of RNA molecules remains resistant for 90 min of incubation. On the gel shown in Figure 2A, it is easy to determine the length of these resistant RNAs by counting the additional nucleotide from the cleavage site. Thus, the shortest substrate that derives from *motB* ends in 3′ with the sequence GGAGUAUAAAA-3′. Although we cannot rule out the possibility that shorter RNA species would be attacked after a much longer period of incubation, they are certainly very poor substrates.

A similar experiment was carried out with the RNA labeled in 3′ (by addition of pCp). As in the preceding experiment, cleavage was more efficient with the longer RNA molecules
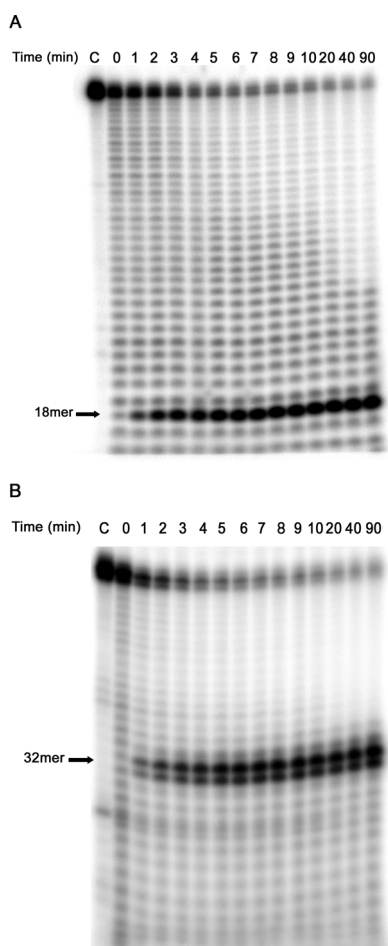
```
                              |
motB        GGUAUAAUGGUCCCGUGGGAGUAUAAAAUCUUUUUAACAAGUGAGAGAUA

motB[denV]  GGUAUAAUGGUCCCGUGGGAGAACACUAUCUUUUUAACAAGUGAGAGAUA

denV[motB]  GGACCUAUCAACUACAGGAGUAUAAAAAAAUGACUCGUAUCAACCUUAC

denV        GGACCUAUCAACUACAGGAGAACACUAAAAUGACUCGUAUCAACCUUAC

                           A   B
                           ↓   ↓
1           CUUAAUUAAAUUUGAGGAGAAACACAUGAAACUAAUCUUUUUAAGCGGUG

1[AU]       CUUAAUUAAAUUUGAGGAGAAUAAAAUGAAACUAAUCUUUUUAAGCGGUG

                           A     B  C
                           ↓     ↓  ↓
1[GG]       CUUAAUUAAAUUUGAGGAGGAGCACAUGAAACUAAUCUUUUUAAGCGGUG
```

Figure 1. Sequences of the *in vitro* synthesized RNAs used in this study. The sequence required for RegB/S1 recognition is in bold letters (see text). The arrows show the sites of cleavage by RegB. The 7 nt sequence that was exchanged between RNAs is underlined.

A

Time (min)   C 0 1 2 3 4 5 6 7 8 9 10 20 40 90

18mer →

B

Time (min)   C 0 1 2 3 4 5 6 7 8 9 10 20 40 90

32mer →

Figure 2. (**A**) Gel electrophoresis of the products of digestion of Na_2CO_3-treated 5′ end-labelled *motB* RNA by RegB in the presence of S1. The time (min) is indicated above the figure. (**B**) Same experiment as in (A) but with the RNA labelled in 3′. Lane C : RNA before Na_2CO_3 treatment.

than with the shorter ones. In this case, resistance to RegB appeared with RNA species that carry two additional nucleotide in 5′ of the cleavage site. These nucleotides are the two consecutive G's of the GGAG motif. Thus, no additional nucleotide in 5′ of the GGAG motif is required in order for RegB/S1 to cleave the *motB* RNA (Figure 2B). The additional strong band, shorter by one nucleotide relative to the

expected 32 nt RNA product (31 + pCp), reflects the fact that the *motB* RNA substrate used in this experiment, although purified by gel electrophoresis, still contained a significant proportion of 48 nt RNA species. In agreement with this result, we found that RegB/S1 could cleave another good RegB substrate, *motA*, truncated such that its sequence starts with the GGAG motif in 5′, albeit with slower kinetics than the entire *motA* RNA (data not shown). This confirms that RNA 5′ of the conserved GGAG sequence is not an absolute requirement for RegB/S1.
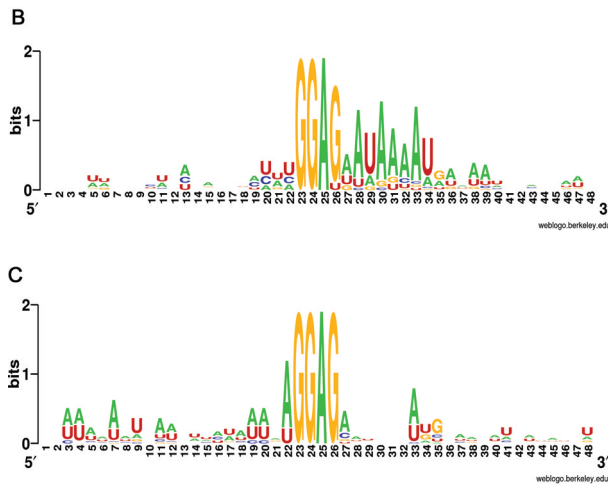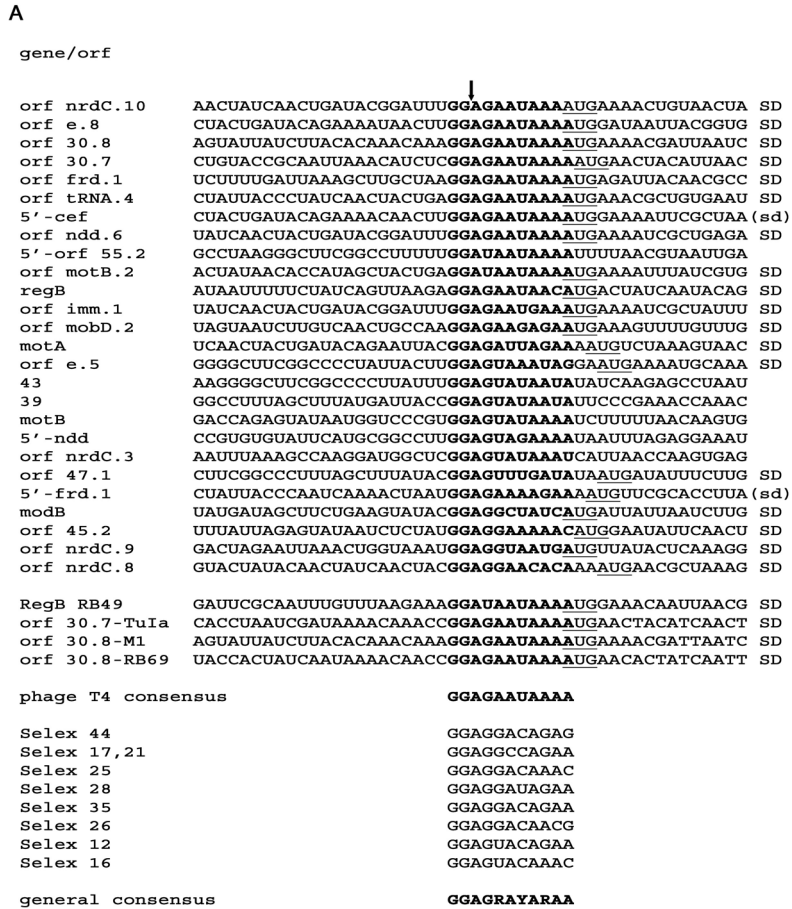
In conclusion, these experiments show that RegB/S1 require additional RNA sequence elements 3′ of GGAG while no specific information seems to be necessary upstream of the GGAG motif. Nevertheless, additional sequences in 5′ and 3′ stimulate the RegB reaction, presumably by facilitating binding of the proteins to the RNA.

T4 transcripts efficiently cleaved by RegB share homologies over 11 nt around the cleavage site

We then asked whether, besides the GGAG motif, any sequence conservation could be found 3′ of this motif among RegB-sensitive T4 RNAs. For this, we identified as many T4 transcripts cleaved by RegB during infection as possible and compared their sequences to those of resistant (or poorly cut) GGAG-containing transcripts. Seventy-one GGAG sequences were found in intergenic regions of T4 early, middle and late transcripts and in non-coding RNAs (tRNAs and introns). Except for three GGAG sequences located in stems of tRNA[pro] and tRNA[ser], they were all tested. Since we found earlier that one GGAU sequence was efficiently cleaved by RegB (7), we also analyzed the 30 intergenic GGAU sequences. RNA isolated at different periods after infection by either T4 wild-type or T4 *regB* R52L mutant phage were analyzed by primer extension. The sites of cleavage were mapped by running an RNA sequence along with the electrophoresis of cDNAs. A faint band or no reverse transcriptase stop at the expected position defined a resistant RNA while the appearance of a strong, RegB-dependent, band at the expected position, defined a good RegB RNA substrate.

Of the 68 tested intergenic GGAG/U motifs, 24 were efficiently cut by RegB while two were moderately cleaved (ORFs nrdC.8 and 47.1) (Figure 3A). Forty-two GGAG sequences were either very poorly cut or fully resistant (data not shown). Only two intergenic GGAU sequences (ORFs 55.2 and *motB*.2) were found to be cut by RegB among the 30 tested sequences (Figure 3A). All the others were resistant (data not shown). Most of the GGAG/U motifs analyzed were Shine–Dalgarno sequences.

The T4 transcripts resistant to RegB were separated into two groups. One of them consisted of sixteen RNAs with an obvious ability to form stable secondary structures that include the GGAG motifs. These include intergenic stem–loop structures, like transcription terminators and introns. Often, the structure contains the stabilizing UUCG tetraloop (data not shown). Whether their resistance arises from this structural feature or because they lack a sequence element is not known. For this reason, we did not consider this group of RNA sequences any further. The GGAG motifs carried by the 26 remaining resistant RNAs cannot be

A

```
gene/orf

orf nrdC.10    AACUAUCAACUGAUACGGAUUUGGAGAAUAAAAAUGAAAACUGUAACUA SD
orf e.8        CUACUGAUACAGAAAAUAACUUGGAGAAUAAAAUGGAUAAUUACGGUG  SD
orf 30.8       AGUAUUAUCUUACACAAACAAAGGAGAAUAAAAAUGAAAACGAUUAAUC SD
orf 30.7       CUGUACCGCAAUUAAACAUCUCGGAGAAUAAAAAUGAACUACAUUAAC  SD
orf frd.1      UCUUUUGAUUAAAGCUUGCUAAGGAGAAUAAAAUGAGAUUACAACGCC  SD
orf tRNA.4     CUAUUACCCUAUCAACUACUGAGGAGAAUAAAAAUGAAACGCUGUGAAU SD
5'-cef         CUACUGAUACAGAAACAACUUGGAGAAUAAAAUGGAAAAUUCGCUAA (sd)
orf ndd.6      UAUCAACUACUGAUACGGAUUUGGAGAAUAAAAUGAAAAUCGCUGAGA  SD
5'-orf 55.2    GCCUAAGGGCUUCGGCCUUUUUGGAUAAUAAAAUUUUAACGUAAUUGA
orf motB.2     ACUAUAACACCAUAGCUACUGGAGAAUAAAAUGAAAAUUUAUCGUG   SD
regB           AUAAUUUUUCUAUCAGUUAAGAGGAGAAUAACAUGACUAUCAAUACAG  SD
orf imm.1      UAUCAACUACUGAUACGGAUUUGGAGAAUGAAAUGAAAAUCGCUAUUU  SD
orf mobD.2     UAGUAAUCUUGUCAACUGCCAAGGAGAAGAGAAUGAAAGUUUUGUUUG  SD
motA           UCAACUACUGAUACAGAAUUACGGAGAUUAGAAAAUGUCUAAAGUAAC  SD
orf e.5        GGGGCUUCGGCCCCUAUUACUUGGAGUAAAUAGGAAUGAAAAAUGCAAA SD
43             AAGGGGCUUCGGCCCCUUAUUUGGAGUAUAAUAUAUCAAGAGCCUAAU
39             GGCCUUUAGCUUUAUGAUUACCGGAGUAUAAUAUUCCCGAAACCAAAC
motB           GACCAGAGUAUAAUGGUCCCGUGGAGUAUAAAAUCUUUUUAACAAGUG
5'-ndd         CCGUGUGUAUUCAUGCGGCCUUGGAGUAGAAAAUAAUUUAGAGGAAAU
orf nrdC.3     AAUUUAAAGCCAAGGAUGGCUCGGAGUAUAAAUCAUUAACCAAGUGAG
orf 47.1       CUUCGGCCCUUUAGCUUUAUACGGAGUUUGAUAUAAUGAUAUUUCUUG SD
5'-frd.1       CUAUUACCCAAUCAAAACUAAUGGAGAAAAGAAAAUGUUCGCACCUUA (sd)
modB           UAUGAUAGCUUCUGAAGUAUACGGAGGCUAUCAUGAUUAUUAAUCUUG SD
orf 45.2       UUUAUUAGAGUAUAAUCUCUAUGGAGGAAAAACAUGGAAUAUUCAACU SD
orf nrdC.9     GACUAGAAUUAAACUGGUAAAUGGAGGUAAUGAUGUUAUACUCAAAGG SD
orf nrdC.8     GUACUAUACAACUAUCAACUACGGAGGAACACAAAAUGAACGCUAAAG SD


RegB RB49      GAUUCGCAAUUUGUUUAAGAAAGGGAUAAUAAAAUGGAAACAAUUAACG SD
orf 30.7-TuIa  CACCUAAUCGAUAAAACAAACCGGAGAAUAAAAAUGAACUACAUCAACT SD
orf 30.8-M1    AGUAUUAUCUUACACAAACAAAGGAGAAUAAAAAUGAAAACGATTAATC SD
orf 30.8-RB69  UACCACUAUCAAUAAAACAACCGGAGAAUAAAAAUGAACACTATCAATT SD


phage T4 consensus       GGAGAAUAAAA


Selex 44                 GGAGGACAGAG
Selex 17,21              GGAGGCCAGAA
Selex 25                 GGAGGACAAAC
Selex 28                 GGAGGAUAGAA
Selex 35                 GGAGGACAGAA
Selex 26                 GGAGGACAACG
Selex 12                 GGAGUACAGAA
Selex 16                 GGAGUACAAAC


general consensus        GGAGRAYARAA
```

B

C

Figure 3. (**A**) Alignment of RNA sequences efficiently cut by RegB from T4 (upper part) or T4-related phages RB49, RB69, TulA and M1 (middle part). Translation initiation codons are underlined. The consensus region is in bold. SD: Shine–Dalgarno sequence. (sd): putative Shine–Dalgarno sequence of very small ORFs. The arrow indicates the site of cleavage by RegB. 5'-followed by the name of a gene or orf means that the sequence shown is located upstream of the TIR of the gene or orf in the leader RNA. Bottom part: alignment of the eight SELEX sequences (23) efficiently cleaved by regB/S1. The new consensus sequence obtained is in bold. R: purine Y: pyrimidine. (**B** and **C**) Sequence logo representation of the efficiently cleaved sequences (B) and resistant sequences (C). The sequences are aligned on the GGAG motif.

included in stable secondary structures. We found that their ability to form structures, as predicted by the M-Fold program (35,36), is not different from that of the efficiently cleaved RNAs shown in Figure 3A. Therefore, their resistance likely reflects the lack of a sequence element.

Comparison of the T4 sequences of the 26 cleaved and 26 resistant, yet unstructured, T4 RNAs showed that the cut RNA sequences contain an A-rich sequence of 7 nt, immediately 3' to the GGAG/U motif, from which C residues are almost completely excluded. The equivalent regions in the

uncut sequences show no particular feature. The strong bias exhibited by the RegB substrates is best illustrated through the Sequence Logo representation (37). Figure 3B shows the results obtained with the 26 well-cleaved RNAs listed in Figure 3A. The score obtained by the nucleotide of the GGA motif is maximal, reflecting their absolute conservation among RegB target sites. This triplet may be the only sequence recognized by RegB (see below). A guanosine in the fourth position (counting from the 5′ most G of the GGAG motif) is also very conserved, but U is possible (2/26). The scores obtained by the 7 nt of the A-rich sequence immediately downstream of -GGAG- are lower, but significantly above that of any nucleotide elsewhere in the 48 nt long segment analyzed. The A's at positions 6, 8 and 11 are the most conserved of the A-rich sequence. In striking contrast, no such good scores are reached by the nucleotide at equivalent positions of the non-cleaved RNA sequences (Figure 3C). The existence of a consensus immediately 3′ of -GGAG- is in agreement with the above finding that cleavage by RegB/S1 requires the presence of a minimum of 7 nt downstream of this motif.

Many T4-related phages code for a homolog of the T4 RegB enzyme (11,38). The sequences of the few cleavage sites identified in four of these phages are shown in the middle part of Figure 3A. In all cases, cleavage occurs in the middle of the GGAG/U motif, which is followed by the A-rich consensus sequence.

Besides the GGAG and GGAU sequences, we tested some intergenic GGAA motifs for their ability to be cleaved by RegB during infection, in particular those that are followed by A-rich sequences. We found that the GGAA sequences are very poorly cut, even when they are followed by a perfect A-rich sequence (data not shown). No intergenic GGAC sequence is followed by an A-rich sequence among the T4 intergenic regions.

Taken together, these observations strongly suggest that the sequence required for efficient RegB cleavage in phage T4 is the 11 nt sequence GGAGAAUAAAA. They also suggest that a GGAU sequence can be cut efficiently provided that it is followed by a strict A-rich consensus sequence. Retrospectively, we note that all 16 T4 RegB resistant, structured RNAs that we put aside (see above) diverge considerably from the consensus determined in this study.

In bacteriophage T4, the A-rich sequence associated with the GGAG sequence is sufficient for RegB/S1 recognition

To assess the importance of the A-rich sequence, we analyzed the RegB/S1 cleavage reaction with chimeric RNAs in which this region was exchanged from a good substrate to a non-substrate and vice versa. All RNAs tested in this study were obtained by *in vitro* transcription. In a first experiment, we analyzed the consequences of replacing the sequence of 7 nt immediately 3′ of -GGAG- in the *motB* RNA by the sequence found at the same place in the *denV* RNA. *In vivo*, the Shine–Dalgarno sequence (GGAG) of the *denV* gene is resistant to RegB (39). In the resulting chimeric *motB[denV]* RNA (Figure 1), the three A's at the most conserved positions (6, 8 and 11) were maintained but four changes were introduced. (i) U to A in position 5. A is
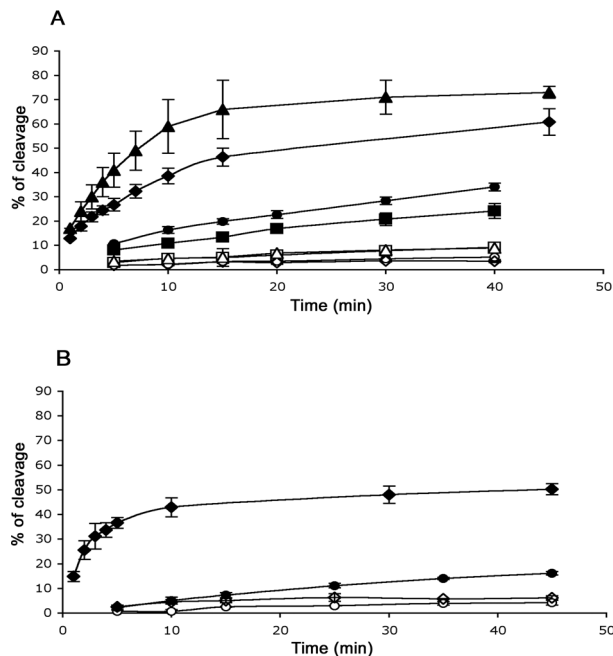
Figure 4. Kinetics of digestion by RegB alone (open symbols) or RegB in the presence of S1 (filled symbol) of *motB* (Δ), *denV* (○), *motB[denV]* (□), *denV[motB]* (◇) (**A**) and gene *1* (○), gene *1[AU]* (◇) RNA (**B**). Experiments were done in triplicate. The conditions are described in Materials & Methods.

favored over any other base in this position and is twice as frequent as U among the well-cleaved T4 RNAs (Figure 3A). (ii) U to C in position 7. No C was found in this position in the cleaved T4 RNAs. (iii) A to C in position 9. No C was found in this position in the cleaved T4 RNAs. (iv) A to U in position 10. In this position, A residues are favored and U's are rare (3/26). In the presence of S1, the reaction with *motB* RNA was fast for the first 10 min and soon afterwards slows down, reaching a plateau when 80% of the initial RNA substrate was cleaved. The apparent forward rate constant (k_{on}) was 0.07/min (Figure 4A and Table 1). With *motB[denV]* RNA, the reaction rate decreased more than 10-fold (k_{on} = 0.006/min) (Figure 4A and Table 1).

We then modified the sequence of *denV* RNA so as to replace 7 nt immediately 3′ of -GGAG- by the A-rich sequence of *motB*. This produced *denV[motB]* chimeric RNA (Figure 1). In the presence of S1, *denV* RNA was rather weakly cleaved by RegB/S1 *in vitro* (k_{on} = 0.015/min). With *denV[motB]* RNA, the cleavage rate was 4 times higher (k_{on} = 0.06/min) (Figure 4A and Table 1).

A similar experiment was conducted with the gene 1 transcript. After infection by T4, this T4 transcript exhibits a unique pattern of cleavage by RegB: It is cut very weakly at two sites (labeled A and B in Figure 1), between G and A, within the Shine–Dalgarno sequence of the gene (7). The sequence 3′ of the GGAG motif contains two C's, in position 8 and 10 (Figure 1). *In vitro*, in the presence of S1, gene 1 RNA was indeed a very poor substrate (k_{on} = 0.004) (Figure 4B and Table 1). However, chimeric RNA 1[AU] in which the 3′ sequence was replaced by the canonical sequence AAUAAAA was cut at the classical position (position B) at a rate more than 30 times higher

(k_{on} = 0.140) (Figure 4B and Table 1). The cut located upstream (site A in Figure 1) in the wild-type sequence was very much weakened (data not shown).

Given that RegB is inhibited by RNA secondary structures (16), it was important to check what consequence the mutations could have had on the structure of the molecules. Possible structures in all six RNAs, *motB*, *denV*, *motB*[*denV*], *denV*[*motB*], *1* and *1*[*AU*] were probed with RNases A, T1, T2 and V1. Slight changes in the reactivity of some regions of these RNAs to the RNases could be observed. However, in no case did the change in sequence lead to a detectable change in structure involving the GGAG motif and surrounding sequences (data not shown). Therefore, the changes in cleavage efficiency observed with the mutants reflect changes in the sequence and not in the structure of the RNAs.

The chimeric RNA *denV*[AU], in which the A-rich consensus sequence AAUAAAA replaced the normal sequence of the *denV* transcript, 3′ of -GGAG-, was recombined into the T4 genome. RNA isolated after T4 wild-type or T4 *denV*[AU] phage infection was analyzed by primer extension. In agreement with the result obtained *in vitro* with the *denV*[*motB*] chimeric RNA, the *denV*[AU] transcript was cut efficiently *in vivo* in the middle of the GGAG sequence, like a natural RegB-sensitive transcript (Figure 5).

Thus, replacing the A-rich sequence of a good substrate by a sequence from a non-substrate considerably decreases the efficiency of cleavage by RegB/S1. Conversely, it is possible to convert a GGAG-carrying resistant RNA into a good substrate for RegB/S1 by grafting the A-rich sequence immediately downstream of the GGAG motif.

Guanine in positions 4 and 5 enhances cleavage by RegB in the presence of S1

As shown above, two GGAU sequences are efficiently cleaved by RegB in phage T4 (orf *55.2* and *motB.2*) and one was found in the T4-related phage, RB69 (11) (Figure 3A). Thus, a U in position 4 seems tolerated, but under strict conditions as the great majority of the intergenic GGAU sequences in T4 are resistant to RegB. To assess the importance of the nature of the nt immediately 3′ to the GGA motif, we analyzed the ability of RegB to cleave, in the presence and absence of S1, *motB*-derived RNAs carrying each of the three possible substitutions in that position. Enzyme kinetics showed that any substitution of the 4th G resulted in a marked decrease of the reaction speed, with the following hierarchy: GGAU > GGAA > GGAC (Figure 6A). The corresponding kinetic constants are shown in Table 1.

RNase probing of the structure of all four RNAs shows that the G to A and G to U changes did not result in the formation of new RNA structures that would hinder the accessibility of the RegB nuclease to its target site (data not shown). Thus, the decreased activity of RegB on these RNAs is a direct consequence of the sequence changes. However, the G to C change strongly affected the accessibility of RNases T1 and A to a significant portion of the molecule, including the GGAC motif, as revealed by the severe weakening of

Table 1. Kinetic constant of RegB cleavage of the RNAs used in this study with or without S1

RNA	k_{on} (+S1)	k_{on} (−S1)	Stimulation factor by S1
motB-G (wild-type)	0.080	0.003	27
motB-U	0.025	0.003	8.3
motB-A	0.009	0.005	1.8
denV	0.015	0.001	15
motB[denV]	0.010	0.003	3.3
denV[motB]	0.060	0.001	60
1	0.004	0.001	4
1[AU]	0.140	0.004	35
1[GG]	0.05	0.002	25

We considered the simple phenomenological model represented by the reaction: S→P. The corresponding equation [A(t) = A$_0$ e^{-kt} + C] was fitted to the experimental curves to determine the apparent forward rate constant (k_{on}). It was not possible to obtain a correct fit between the model and all data points. In these cases, only the first points from 1 to 15 min were taken into account.
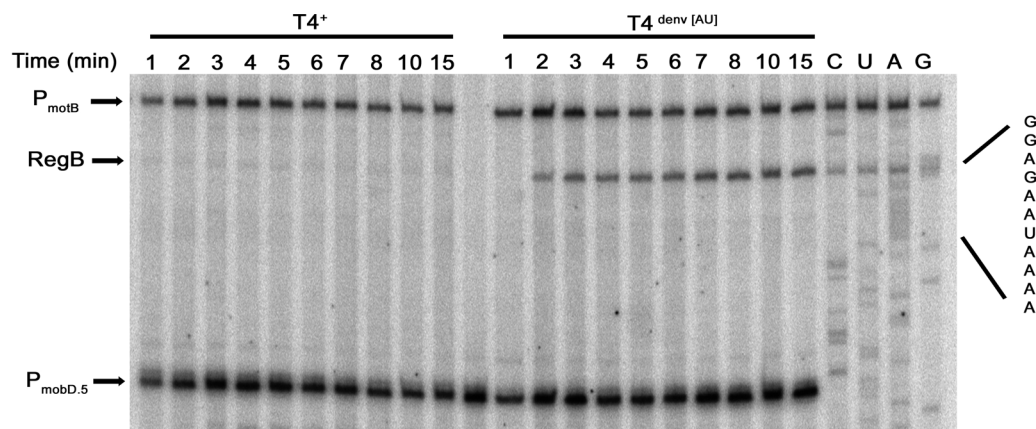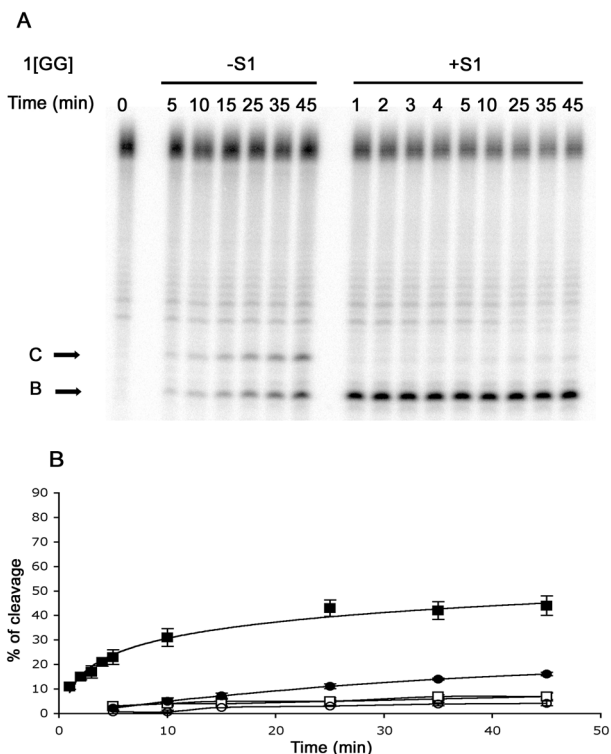

Figure 5. Analysis of the susceptibility of *denV* and *denV[AU]* mRNAs to RegB after T4 infection. Samples of T4 wild-type or T4 *denV[AU]* infected *E.coli* B^E were withdrawn at the times indicated above the figure for immediate RNA extraction. The RegB cut was probed by primer extension using the denV-25 oligonucleotide. cDNAs were separated by electrophoresis on a 7% polyacrylamide–urea gel along with the sequence of the *denV[AU]* RNA (right). RNA to be sequenced was extracted 3 min after infection by T4 *denV[AU]*. Arrows indicate the reverse transcriptase stops corresponding to the *motB* and *modD.5* promoters. The *mobD.5* transcript is not sensitive to RegB and was used as an internal control of infection and RNA extraction. The position of RegB cleavage is also shown.
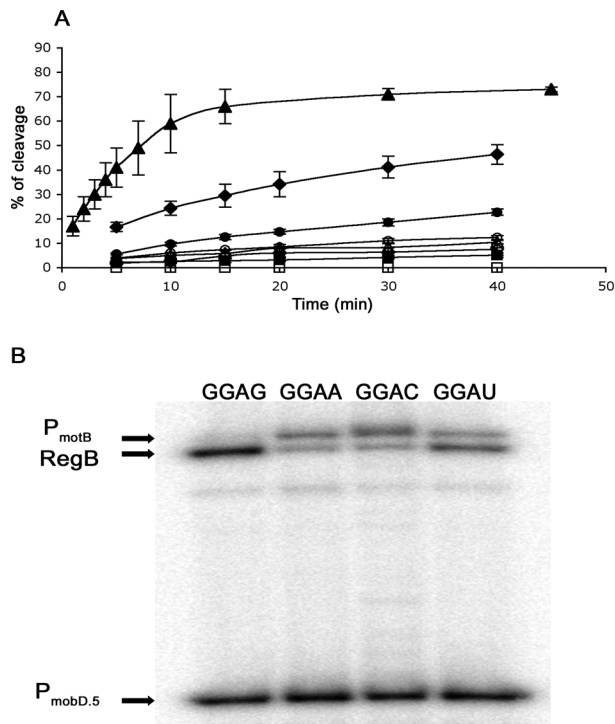
Figure 6. (**A**) Kinetics of digestion by RegB of *motB* RNA derivatives carrying a substitution in the fourth position of the GGAG motif. Open symbols: kinetics without S1. Filled symbols: Kinetics in the presence of S1 *motB*-GGAG (wild-type) (△), *motB*-GGAA (○), *motB*-GGAC (□), and *motB*-GGAU (◇). Experiments were done in triplicate. (**B**) Analysis of the susceptibility of *motB*-GGAG/A/C/U mRNAs to RegB after T4 infection. Samples of T4 wild-type or T4 GGAA/C/U infected *E.coli* B^E were withdrawn at 3 min for immediate RNA extraction. The RegB cut was probed by primer extension using the motB oligonucleotide. cDNAs were separated by electrophoresis on a 7% polyacrylamide–urea gel.

Figure 7. (**A**) Gel electrophoresis of the kinetics of digestion of *1[GG]* RNA by RegB with or without S1. B, C: products of digestion by RegB corresponding to the cleavages shown in Figure 1. (**B**) Kinetics of digestion by RegB (open symbols) or RegB/S1 (filled symbols) of gene *1* (○) and gene *1[GG]* (□) RNA. Experiments were done in triplicate.

several bands (data not shown). Thus, under our *in vitro* conditions, the *motB*-GGAC RNA folded into a rather stable structure in which the GGAC nucleotide were perfectly paired. This probably explains its total resistance to RegB (see below).

The susceptibility to RegB of transcripts carrying these three mutations in *motB* was also examined *in vivo*. The *motB* gene and surrounding regions can be deleted without affecting phage growth (40). This ensures that any mutation introduced within and around the GGAG sequence can be selected and studied *in vivo*. The three mutations were transferred into the T4 genome by homologous recombination. RNA isolated 3 min after infection by wild-type and mutant phages was analyzed by primer extension. In agreement with our *in vitro* results, the efficiency of cleavage of the mutated transcripts was substantially reduced and the same hierarchy among the mutations was observed (Figure 6B). However, in this case the GGAC transcript was cleaved to a significant extent, even though it was the least affected by RegB. This indicates either that the structure that forms *in vitro* is counteracted *in vivo*, or that the cleavage by RegB occurs before it can form. In conclusion, in the context of *motB* RNA, the G in position 4 is by far preferred over any other nt, and this result seems general as assessed by the frequency of G in

this position *in vivo* (Figure 3A). However, within this context, other nucleotide are tolerated, with U being preferred over A or C.

The above conclusion that the recognition site for RegB/S1 is the 11 nt sequence whose consensus is GGAGAAUAAAA suffers from the existence of a famous counter-example. The RNA initially used to monitor RegB purification and to characterize RegB biochemical properties was a decamer whose sequence is: 5′-CUUUGGAGGG (15). Cleavage occurs at the expected position and is stimulated by S1 about 80-fold (15,16). This small RNA contains the GGAGG motif, as do four of the well-cleaved T4 transcripts (Figure 3A) and half of the RNA molecules obtained by SELEX (23). Thus the G in position 5 (in the GGAGG sequence) may have a strong positive effect on the cleavage reaction.

To test this hypothesis, we modified the poorly cleaved gene *1* RNA, replacing the A in position 5 by a G, so as to obtain -GGAGGA- (gene 1[GG] RNA; Figure 1). The ability of the mutated RNA to be cut by RegB in the presence and absence of S1 was analyzed. Figure 7A and B show that the simple presence of a G in position 5 converted the very poor gene 1 RNA substrate into a rather good one. In the presence of S1, the increase in the kinetic constant with respect to wild-type gene 1 RNA was one order of magnitude (k_{on} for gene 1[GG] was 0.05; Table 1).

Interestingly, in the absence of S1, gene 1[GG] RNA was cut twice, between G and A, in the GGAGGA sequence, suggesting that GGA is the motif recognized by RegB.

In the presence of S1, RegB only cleaves the B site efficiently (Figures 1 and 7A).

In our kinetic experiments, we observed that the level of the plateaus varied from one RNA to another, with a maximum cleavage of 80% for *motB* and 50% for gene *1* RNA (Figures 4A, B, 6A and 7B), These maxima depend on the RNA sequence and are unchanged whether we heat the RNA and chill it suddenly on ice or decrease the temperature slowly after heating. We propose that RegB is inhibited by the RNA products. This phenomenon, already observed by others (15,16,26), was not investigated further.

Structure of the RegB/S1 target site

The above result suggests that, due to selection pressure operative *in vivo*, the consensus sequence for RegB/S1 found in phage T4 is a particular arrangement of a looser consensus. In order to enlarge the pool of sequences that would allow us to determine the general structure of the RegB target site, we turned to the list of RNA sequences cleaved by RegB in the presence of S1 obtained by SELEX (23). Unfortunately, not all these RNA sequences could be taken into account since their cleavage rates differ by a factor of at least 20. Our main criterion for a good RegB substrate is its ability to be cleaved efficiently by RegB during phage infection. As shown here [see also (14)], fully resistant T4 RNAs *in vivo* can be cut to a certain extent *in vitro*, depending on the RegB/RNA ratio. Thus, sequences that we would have regarded as bad substrates might well be represented in Jayasena *et al*.'s list. Therefore, we chose to consider only the SELEX sequences for which a good kinetic constant was found. In Figure 3A (lower part), we aligned the subset of well-cleaved SELEX sequences with the good T4 substrates. Only the 11 nt that span the RegB/S1 target site as defined in this study were considered. Although efficiently cut by RegB, the SELEX # 22 RNA sequence was removed from the list because its cleavage does not depend on S1 (16). SELEX # 12 and 26 (in fact, shorter derivatives from which the constant sequences were removed) were shown by Lebars *et al*. (16) to be good RegB substrates. The result was the consensus shown in the lower part of Figure 3A.

In order to evaluate the degree of divergence between the subset of good SELEX substrates chosen and the whole SELEX sequences, we calculated the frequency of occurrence of each nucleotide, in the 11 nt window, for 29 SELEX sequences (five sequences of RNA that Jayasena *et al*. found extremely poorly cut were excluded). The same consensus sequence for RegB/S1 as that shown in Figure 3A (lower part) was found. This analysis allowed comparison of the 26 T4 sequences (Figure 3A) with the 29 SELEX sequences (23). Besides the common GGAG motif, the following analogies and differences could be noticed.

Position 5: A purine was twice as frequent as U in both the SELEX and T4 sequences. However, A is preferred in T4 and G in the SELEX sequences. C is counter-selected in both groups (1/29 SELEX sequences and none in T4).

Position 6: A is strongly preferred in both groups.

Position 7: A pyrimidine is strongly preferred in both groups of sequences, with a bias in favor of C in the SELEX RNAs and in favor of U in the T4 group, where no C at all was found.

Position 8: Strong bias in favor of A in both groups of sequences.

Position 9: Almost total bias in favor of a purine in both groups. A and G are equally probable in the SELEX sequences while A is strongly preferred in T4. No C at all was found and U's were extremely rare in either group of sequences.

Positions 10 and 11: Large majority of A's in both groups of sequences.

These observations show very good but not total agreement between the two groups of sequences, indicating that the two types of evolution, *in vivo* and *in vitro*, have been subjected to different selection pressure.

Activation by S1 depends on the presence of specific 3′ nt of GGA

How does the S1 protein enhance the RegB reaction and what RNA sequence does it require? Two models are possible. In the first, RegB alone would be able to discriminate the good from the bad substrates, and the S1 ribosomal protein would enhance the efficiency of the reaction equally, whatever the RNA considered. Alternatively, RegB alone would not be able to discriminate among the various GGAG-carrying RNAs. It would simply recognize, with low efficiency, the much conserved GGA trinucleotide sequence, irrespective of the context. The S1 protein would play a major role in the discrimination process, presumably through specific interactions with the 11 nt conserved sequence or a part of it.

The first hypothesis predicts that *in vitro*, RegB alone should cleave the good substrates more efficiently than the poor ones. This is clearly not the case. In the absence of S1, RegB, under our standard conditions, is as inefficient on RNAs identified as good substrates *in vivo* as on non-substrates (Figures 4A, B, 6A and 7B; Table 1). Increasing the enzyme concentration did not allow RegB to discriminate better between good and bad substrates (data not shown). In fact, our results favor the second hypothesis. We found that while RegB cleavage of *motB* wild-type RNA was stimulated by S1 27-fold, the *motB*[*denV*] derivative, in which the A-rich sequence comes from a poorer substrate, was stimulated only 3.3-fold (Table 1). Conversely, the stimulation by S1 of *denV* RNA cleavage increased from 15- to 60-fold when the A-rich sequence from *motB* was grafted within the *denV* sequence (*denV*[*motB*] RNA) (Figure 4A and Table 1). Likewise, the 27-fold stimulation observed with *motB* RNA dropped to only 8.3- and 1.8-fold when the fourth G was changed to U and A, respectively (Figure 6A and Table 1). With wild-type gene 1 RNA, S1 stimulated about 4-fold (Table 1). This factor was dramatically increased to 35-fold when the A-rich sequence replaced the wild-type sequence 3′ of the GGAG motif (1[AU] RNA) and to 25-fold with G in position 5 (gene 1[GG] RNA) (Figures 4B and 7B; Table 1). Thus, the changes in RegB efficiency consecutive to alterations of the sequence downstream of -GGA- directly reflect changes in the ability of S1 to stimulate the cleavage reaction. G residues in positions 4 and 5 play a major role in the stimulation by the S1 co-factor, as well as some other nucleotide within the A-rich sequence. This work shows that the S1 protein plays a crucial role in the selection of RegB substrates.

DISCUSSION

The work reported here shows that the T4 RegB endoribonuclease target site spans 11 nt of single-stranded RNA carrying the 100% conserved GGA triplet at the 5′ end and a degenerate consensus sequence immediately downstream. Progressive deletion analysis, alignment of well-cleaved and resistant T4 RNA sequences and nucleotide changes at specific positions led to the following consensus: GGAGRA-YARAA, where R is a purine and Y a pyrimidine. The part of the site 3′ to the very conserved GGAG motif, although degenerate, exhibits strong constraints: extreme rarity of C residues in positions 5 and 6; total absence of C in position 9; strong bias in favor of A residues in positions 6, 8, 10 and 11; strong bias in favor of a pyrimidine in position 7. The presence of a C in position 9 of the *denV* sequence likely accounts for the poor performance of RegB/S1 on this RNA and its derivative, *motB*[*denV*]. The almost totally resistant gene 1 RNA diverges at three positions with respect to the consensus we defined: an A instead of a pyrimidine in position 7 and the presence of cytosines instead of A residues in positions 8 and 10.

Our data support the notion that RegB alone recognizes only the trinucleotide GGA, which it cleaves very inefficiently. Indeed, (i) whether *in vivo* or *in vitro*, all RegB cleavage occurs within the GGA sequence, whatever the nucleotide context (Figures 3A, 4B, 6A, B and 7A). The only exception is the very weak cut found within the first G-A link in the GAGGAG sequence of T4 gene 1 (Figure 1). (ii) RegB alone is able to cleave equally the two GGA sequences in tandem in the gene 1[GG] RNA (Figures 1 and 7B). (iii) The GGA sequence was found in every RNA cleaved by RegB in the presence of S1 obtained by SELEX (23).

Our results also show that stimulation of RegB activity by the S1 ribosomal protein depends at least on the nucleotide immediately 3′ of -GGA-. In fact, we found that changes introduced within this sequence have no consequence on the rate of cleavage by RegB alone, which always remains low, but affect the degree of RegB stimulation by S1 (Table 1). The fact that the only 2 (of 26) well-cleaved RNA sequences that have a U immediately 3′ of GGA are followed by strict consensus sequences, suggests that the deleterious effect of U in that position can be compensated for only by a perfect A-rich sequence. These results strongly suggest that the S1 protein acts by interacting specifically at least with the nucleotide 3′ of GGA in the RegB target sites. They also show that S1 plays a major role in the selection of RegB substrates.

Does this latter result mean that the RegB target site is essentially an S1 binding site? Remarkably, the length of the RegB/S1 site precisely matches the size of 10–12 nt required for binding of one molecule of S1 to RNA (17). This suggests that the whole sequence, including the GGA motif, interacts with S1. Furthermore, the GGA motif, the A-rich region and the ACA motif found in SELEX sequences are present in other S1 target sites: A-rich or AU-rich sequences are found in translational enhancers, in mRNA poly(A) tails (41,42) and in the translational operator of the *rpsA* gene that codes for ribosomal protein S1 (43). The *rpsA* mRNA leader also contains two GGA sequences.

The GGA sequence as well as the ACA triplet are conserved sequences found in the loops of pseudoknots obtained by SELEX carried out against S1 (44). The ACA motif is also found in some translational enhancer sequences (41,45).

We propose that these different motifs interact specifically with different sites on the S1 protein. We further propose that if more than one of these motifs lie within the distance covered by one molecule of S1, their binding to S1 may induce constraints on the RNA, depending on their relative position on the RNA. The RegB target site contains two of these specific signals: -GGA- and, in 3′, the A-rich sequence (with -ACA- frequently found among the SELEX sequences). We suggest that upon binding to S1, the RegB target site undergoes constraints, especially upon GGA, that are exploited by RegB to act efficiently. Thus, the discriminating role of S1 would not be based on its ability to bind or not to bind (or bind less strongly) RNA substrates and non-substrates, respectively, but to induce or not the conformational constraint on the GGA sequence which is the true RegB target. In fact, most of the intergenic RegB-resistant GGAG motifs are located in TIRs of highly translated mRNAs (coding for capsid proteins). There is little doubt that S1 interacts with these regions quite well. For instance, we found, by gel retardation assays, that S1 binds equally well to RegB substrates and to resistant RNAs, both carrying translation initiation regions (TIRs) with -GGAG- as Shine–Dalgarno sequence (S. Durand and M. Uzan, unpublished data). These data suggest that S1 accelerates the reaction not by increasing RNA-binding but by enhancing activity of the enzyme. Experiments using homopolymers led to the conclusion that protein S1 has more than one site for binding nucleic acids (17,42). Here, we assume that this holds true for nucleic acids of higher complexity as well. Our model is supported by the finding that binding of S1 to the AU-rich sequences of the *rpsA* 5′-untranslated regions (5′-UTRs) leads to enhanced exposure of two GGA sequences located nearby (43). In addition, we found previously that a particular structure is necessary to permit a small RNA (SELEX #22) to be efficiently cleaved by RegB in the absence of the S1 co-factor (16). In this structure, the GGA sequence is located in a loop, while the two nucleotide immediately adjacent in 3′ (GC) are base paired. We speculated that this conformation caused the -GGA- to be constrained in such a way that the G-A link would be particularly exposed, thus facilitating RegB attack. The importance, for S1 stimulation, of these G residues (this work) and the strong stimulation by S1 of JR10 RNA, while this RNA has only three G's after the GGA motif (15,16), are in favor of this hypothesis.

The comparison of the most frequent T4 RegB sites to those obtained through the SELEX experiment suggests that the T4 consensus, GGAGAAUAAAA, is a particular nucleotide combination of a looser consensus. The differences observed between *in vivo* and *in vitro* evolution could have several causes. In prokaryotes, genetic information is very compact, so that *in vivo*, several constraints are expected to be superimposed on those for RegB/S1. One of these is the base composition of the T4 genome which is 65.5% AT (46). This bias is compatible with some characteristics of the T4 sites, like the strong preference for A rather than an equal occurrence of A and G in positions 5 and 9, and the strong preference for U rather than C in position 7. Another

constraint may arise from the fact that most RegB target sites are in TIRs. This location could explain in part the rare occurrence of the GGAGG motif in phage T4 (4/26), while half of the SELEX RNAs have this pentanucleotide. In terms of energetics, GGAGG is a better Shine–Dalgarno sequence than GGAG since it provides an additional GC bp with the 16S rRNA. However, too long SD sequences may not be the best in terms of translation efficiency, since stronger binding of the 30S ribosomal subunit to mRNA can impair the escape of the ribosome (41).

The detailed mechanism of action of RegB may also differ *in vitro* and during phage infection. We found in this work that, in spite of the fact that the *motB*-GGAC mutant RNA is totally resistant to RegB *in vitro*, presumably as a consequence of its ability to form a stable secondary structure, this same G to C mutation does not lead to RegB resistance *in vivo* but only to reduced cleavage efficiency (Figure 7A and B). A similar situation was encountered with *modB* T4 RNA (Figure 3A and S. Durand and M. Uzan, unpublished data). One possibility is that *in vivo*, cleavage occurs on the nascent RNA, before the structure can form. Taking into account the tight coupling between transcription and translation in prokaryotes, these observations support a model in which during infection, RegB uses the 30S ribosomal subunit bound to nascent mRNAs as a source of S1 protein, rather than the free protein (13). The occurrence of RegB cuts earlier than 45 s post-infection (9), also suggests that RegB activity *in vivo* is facilitated by another process related to transcription and/or translation. Several *E.coli* endoribonucleases that belong to toxin–antitoxin systems and are induced under nutritional stress conditions (47,48) present mechanistic analogies with RegB. The YoeB and RelE nucleases act on mRNA engaged in translation. MazF and RelE are also specific for trinucleotide sequences. The RelE enzyme and in some cases, YoeB, like RegB, cleave their target mRNAs between the second and third nucleotide. RelE cleavages occur within the codon engaged in the 'A' site of the ribosome, but not in the free RNA molecule (49,50). Finally, in the case of RegB, the S1 protein might well be provided by the transcriptional apparatus. Indeed, recent reports support the view that S1 is in permanent interaction with nascent mRNA, possibly through its indirect binding to RNA polymerase (51,52).

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Steege,D.A. (2000) Emerging features of mRNA decay in bacteria. *RNA*, **6**, 1079–1090.
2. Regnier,P. and Arraiano,C.M. (2000) Degradation of mRNA in bacteria: emergence of ubiquitous features. *Bioessays*, **22**, 235–244.
3. Dreyfus,M. and Regnier,P. (2002) The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell*, **111**, 611–613.
4. Dodson,R.E. and Shapiro,D.J. (2002) Regulation of pathways of mRNA destabilization and stabilization. *Prog. Nucleic Acid Res. Mol. Biol.*, **72**, 129–164.
5. Coburn,G.A. and Mackie,G.A. (1999) Degradation of mRNA in *Escherichia coli*: an old problem with some new twists. *Prog. Nucleic Acid Res. Mol. Biol.*, **62**, 55–108.
6. Nicholson,A.W. (1999) Function, mechanism and regulation of bacterial ribonucleases. *FEMS. Microbiol. Rev.*, **23**, 371–390.
7. Sanson,B. and Uzan,M. (1995) Post-transcriptional controls in bacteriophage T4: roles of the sequence-specific endoribonuclease RegB. *FEMS. Microbiol. Rev.*, **17**, 141–150.
8. Uzan,M. (2001) Bacteriophage T4 RegB endoribonuclease. *Meth. Enzymol.*, **342**, 467–480.
9. Uzan,M., Favre,R. and Brody,E. (1988) A nuclease that cuts specifically in the ribosome binding site of some T4 mRNAs. *Proc. Natl Acad. Sci. USA*, **85**, 8895–8899.
10. Saida,F., Uzan,M. and Bontems,F. (2003) The phage T4 restriction endoribonuclease RegB: a cyclizing enzyme that requires two histidines to be fully active. *Nucleic Acids Res.*, **31**, 2751–2758.
11. Piesiniene,L., Truncaite,L., Zajanckauskaite,A. and Nivinskas,R. (2004) The sequences and activities of RegB endoribonucleases of T4-related bacteriophages. *Nucleic Acids Res.*, **32**, 5582–5595.
12. Repoila,F., Tetart,F., Bouet,J.Y. and Krisch,H.M. (1994) Genomic polymorphism in the T-even bacteriophages. *EMBO J.*, **13**, 4181–4192.
13. Sanson,B. and Uzan,M. (1993) Dual role of the sequence-specific bacteriophage T4 endoribonuclease RegB. mRNA. inactivation. and. mRNA. destabilization. *J. Mol. Biol.*, **233**, 429–446.
14. Sanson,B., Hu,R.M., Troitskayadagger,E., Mathy,N. and Uzan,M. (2000) Endoribonuclease RegB from bacteriophage T4 is necessary for the degradation of early but not middle or late mRNAs. *J. Mol. Biol.*, **297**, 1063–1074.
15. Ruckman,J., Ringquist,S., Brody,E. and Gold,L. (1994) The bacteriophage T4 regB ribonuclease. Stimulation of the purified enzyme by ribosomal protein S1. *J. Biol. Chem.*, **269**, 26655–26662.
16. Lebars,I., Hu,R.M., Lallemand,J.Y., Uzan,M. and Bontems,F. (2001) Role of the substrate conformation and of the S1 protein in the cleavage efficiency of the T4 endoribonuclease RegB. *J. Biol. Chem.*, **276**, 13264–13272.
17. Subramanian,A.R. (1983) Structure and functions of ribosomal protein S1. *Prog. Nucleic Acid Res. Mol. Biol.*, **28**, 101–142.
18. Sorensen,M.A., Fricke,J. and Pedersen,S. (1998) Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in *Escherichia coli in vivo*. *J. Mol. Biol.*, **280**, 561–569.
19. Bycroft,M., Hubbard,T.J., Proctor,M., Freund,S.M. and Murzin,A.G. (1997) The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell*, **88**, 235–242.
20. Schubert,M., Edge,R.E., Lario,P., Cook,M.A., Strynadka,N.C., Mackie,G.A. and McIntosh,L.P. (2004) Structural characterization of the RNase E S1 domain and identification of its oligonucleotide-binding and dimerization interfaces. *J. Mol. Biol.*, **341**, 37–54.
21. Buttner,K., Wenig,K. and Hopfner,K.P. (2005) Structural framework for the mechanism of archaeal exosomes in RNA processing. *Mol. Cell*, **20**, 461–471.
22. Mitchell,P. and Tollervey,D. (2000) Musing on the structural organization of the exosome complex. *Nature Struct. Biol.*, **7**, 843–846.
23. Jayasena,V.K., Brown,D., Shtatland,T. and Gold,L. (1996) *In vitro* selection of RNA specifically cleaved by bacteriophage T4 RegB endonuclease. *Biochemistry*, **35**, 2349–2356.
24. Bear,D.G., Ng,R., Van,D.D., Johnson,N.P., Thomas,G., Schleich,T. and Noller,H.F. (1976) Alteration of polynucleotide secondary structure by ribosomal protein S1. *Proc. Natl Acad. Sci. USA*, **73**, 1824–1828.
25. Szer,W., Hermoso,J.M. and Boublik,M. (1976) Destabilization of the secondary structure of RNA by ribosomal protein S1 from *Escherichia coli*. *Biochem. Biophys. Res. Commun.*, **70**, 957–964.

26. Bisaglia,M., Laalami,S., Uzan,M. and Bontems,F. (2003) Activation of the RegB endoribonuclease by the S1 ribosomal protein is due to cooperation between the S1 four C-terminal modules in a substrate-dependant manner. *J. Biol. Chem.*, **278**, 15261–15271.

27. Kreuzer,K.N. and Selick,H.E. (1994) Directed Insertion/Substitution mutagenesis. In Karam,J.D. (ed.), *Molecular Biology of Bacteriophage T4*. ASM Press, Washington D.C, pp. 452–454.

28. Studier,F.W. and Moffatt,B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, **189**, 113–130.

29. Dennert,G. and Henning,U. (1968) Tyrosine-incorporating amber suppressors in *Escherichia coli* K12. *J. Mol. Biol.*, **33**, 327–329.

30. Neidhardt,F.C., Bloch,P.L. and Smith,D.F. (1974) Culture medium for enterobacteria. *J. Bacteriol.*, **119**, 736–747.

31. Wower,I.K., Zwieb,C.W., Guven,S.A. and Wower,J. (2000) Binding and cross-linking of tmRNA to ribosomal protein S1, on and off the *Escherichia coli* ribosome. *EMBO J.*, **19**, 6612–6621.

32. Zawadzki,V. and Gross,H.J. (1991) Rapid and simple purification of T7 RNA polymerase. *Nucleic Acids Res.*, **19**, 1948.

33. Abramoff,M.D., Magelhaes,P.J. and Ram,S.J. (2004) Image Processing with ImageJ. *Biophotonics Int.*, **11**, 36–42.

34. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual, 2 edn.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

35. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

36. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.

37. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.

38. Truncaite,L., Zajanckauskaite,A., Arlauskas,A. and Nivinskas,R. (2006) Transcription and RNA processing during expression of genes preceding DNA ligase gene 30 in T4-related bacteriophages. *Virology*, **344**, 378–390.

39. Uzan,M., Brody,E. and Favre,R. (1990) Nucleotide sequence and control of transcription of the bacteriophage T4 *motA* regulatory gene. *Mol. Microbiol.*, **4**, 1487–1496.

40. Homyk,T., Jr and Weil,J. (1974) Deletion analysis of two nonessential regions of the T4 genome. *Virology*, **61**, 505–523.

41. Komarova,A.V., Tchufistova,L.S., Supina,E.V. and Boni,I.V. (2002) Protein S1 counteracts the inhibitory effect of the extended Shine–Dalgarno sequence on translation. *RNA*, **8**, 1137–1147.

42. Kalapos,M.P., Paulus,H. and Sarkar,N. (1997) Identification of ribosomal protein S1 as a poly(A) binding protein in *Escherichia coli*. *Biochimie*, **79**, 493–502.

43. Boni,I.V., Artamonova,V.S., Tzareva,N.V. and Dreyfus,M. (2001) Non-canonical mechanism for translational control in bacteria: synthesis of ribosomal protein S1. *EMBO J.*, **20**, 4222–4232.

44. Ringquist,S., 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: comparison of natural and unnatural binding sites. *Biochemistry*, **34**, 3640–3648.

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

46. Miller,E.S., Kutter,E., Mosig,G., Arisaka,F., Kunisawa,T. and Ruger,W. (2003) Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.*, **67**, 86–156.

47. Gerdes,K., Christensen,S.K. and Lobner-Olesen,A. (2005) Prokaryotic toxin-antitoxin stress response loci. *Nature Rev. Microbiol.*, **3**, 371–382.

48. Condon,C. (2006) Shutdown decay of mRNA. *Mol. Microbiol.*, **61**, 573–583.

49. Pedersen,K., Zavialov,A.V., Pavlov,M.Y., Elf,J., Gerdes,K. and Ehrenberg,M. (2003) The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell*, **112**, 131–140.

50. Christensen,S.K., Maenhaut-Michel,G., Mine,N., Gottesman,S., Gerdes,K. and Van Melderen,L. (2004) Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: involvement of the yefM-yoeB toxin-antitoxin system. *Mol. Microbiol.*, **51**, 1705–1717.

51. Sukhodolets,M.V. and Garges,S. (2003) Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry*, **42**, 8022–8034.

52. Sukhodolets,M.V., Garges,S. and Adhya,S. (2006) Ribosomal protein S1 promotes transcriptional cycling. *RNA*, **12**, 1505–1513.