The EMBO Journal vol.15 no.16 pp.4402–4413, 1996

A long-range RNA–RNA interaction forms a pseudoknot required for translational control of the IF3–L35–L20 ribosomal protein operon in *Escherichia coli*

C.Chiaruttini[1], M.Milet and M.Springer

UPR 9073, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France

[1]Corresponding author

In the IF3–L35–L20 operon encoding translation initiation factor 3 (IF3) and the two ribosomal proteins L35 and L20, the expression of the genes that code for the two ribosomal proteins is negatively regulated at the translational level by the cellular concentration of L20. This translational repressor directly regulates the expression of the gene encoding L35 and, via translational coupling, that of its own gene. Mutations that affect the control of the L35 gene were found exclusively at two sites: the first is located ~300 nucleotides upstream, and the second immediately 5′ of the translation initiation site of the L35 gene. Mutations that fall between these two sites have little or no effect on the control, and the lack of effect of a deletion in the intervening region confirms this finding. RNA structure mapping *in vitro* suggests that the first site pairs with the second. We show that this pairing is also likely to occur *in vivo* because single mutations in either of these sites affect control, but base pair compensatory mutations re-establish control. We propose that these two distant sites can base-pair to form a long-range pseudoknot which is required for the control of the expression of the L35 gene.

Keywords: feedback regulation/ribosomes/tertiary RNA interaction/translation

Introduction

The biological functions performed by RNA molecules depend, as for any other biological molecules, on their three-dimensional structure. The functional domains of RNAs are built up by secondary and tertiary interactions which allow specific motifs to be properly oriented in space. These motifs can, in turn, be responsible for either long-range RNA–RNA interactions and/or recognition by specific proteins. Proper RNA folding and specific protein–RNA interactions are known to be essential to many biological processes such as ribosome assembly and function (for review, see Gutell, 1993), splicing of group I and group II introns (Burke *et al.*, 1987; Jacquier and Michel, 1987; Michel and Westhof, 1990), nuclear pre-mRNA splicing (Gallinaro *et al.*, 1994), cleavage of precursor tRNAs by RNase P (Haas *et al.*, 1991) and control of gene expression at both the transcriptional (Landick and Yanofsky, 1987; Das, 1993) and translational level (Zengel and Lindahl, 1994; Springer, 1996).

Regulation at the translational level in prokaryotes is often caused by the binding of a regulatory protein to a specific region of the mRNA. The attached repressor decreases translation either by directly inhibiting the binding of the ribosome (Winter *et al.*, 1987; Hartz *et al.*, 1989; Moine *et al.*, 1990) or by trapping it in an inactive initiation complex (Philippe *et al.*, 1993; Spedding *et al.*, 1993). The region of mRNA responsible for the control is called a translational operator. Quite a few prokaryotic operators have been studied in detail using a combination of genetic and biochemical approaches. There are no strict rules as far as the size and structure of translational operators are concerned. Among the smallest operators is that of the bacteriophage T4 gene 44, whose expression is repressed by the RegA protein, and which is ~15 nucleotides long and apparently unstructured (Miller *et al.*, 1994). The best studied of all the translational operators, in terms of repressor binding, is that of the replicase gene of RNA coliphage R17 (Witherell *et al.*, 1991). Its expression is repressed by the coat protein and its operator consists of a 21 nucleotide-long stem–loop structure with a single nucleotide bulge in the stem. More typical translational operators are ~100 nucleotides long and consist of several stem–loop structures (containing bulges and loops of various sizes) separated by single-stranded regions. Several of these translational operators form RNA pseudoknots that have been shown to be essential for control (Tang and Draper, 1989; Philippe *et al.*, 1990).

However, much larger operators of ~200–400 nucleotides have been reported in a few cases such as the *Escherichia coli* L10–L7/L12 or IF3–L35–L20 ribosomal protein operons (Friesen *et al.*, 1983; Lesage *et al.*, 1990). The expression of the L10–L7/L12 operon is negatively regulated by the cellular concentration of L10 (Friesen *et al.*, 1983; Christensen *et al.*, 1984) which binds 140 nucleotides upstream of the first gene of the operon. Regulation at a distance (at the translation initiation site of the first cistron, >100 nucleotides downstream of the repressor binding site) was proposed to occur through a long-range conformational change: the binding of the repressor induces a conformational change that 'propagates' by a set of alternative secondary structures to the translation initiation region of the first cistron and results in the sequestering of its Shine–Dalgarno (SD) sequence.

The present work deals with the IF3–L35–L20 operon. IF3 is a translation initiation factor encoded by the *infC* gene and L35 and L20, two ribosomal proteins, respectively encoded by *rpmI* and *rplT*. These genes are transcribed from four promoters (Wertheimer *et al.*, 1988; Lesage *et al.*, 1990; see Figure 1). The first, p_{thrS}, is located upstream of *thrS*, the gene encoding threonyl-tRNA synthetase. The second (p_0) and third ($p_{0'}$) promoters are located within *thrS*, and the fourth (p_1) within *infC*. The major promoters of the operon are p_{thrS} and $p_{0'}$. The operon also contains an internal transcriptional terminator

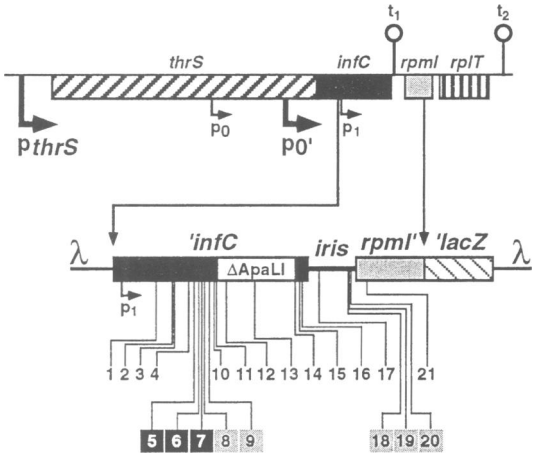

Fig. 1. Location of the mutations and ΔApaLI deletion boundaries and their effect on the repression of an *rpmI'–'lacZ* fusion. The upper part of the figure shows the structure of the *E.coli* chromosome at 38 min with *thrS*, *infC*, *rpmI* and *rplT* coding for threonyl-tRNA synthetase, translation initiation factor IF3 and the ribosomal proteins, L35 and L20, respectively. The p_{thrS}, p_0, $p_{0'}$, p_1 promoters and the two ρ-independent transcriptional terminators, t_1 and t_2, are represented by arrows indicating the direction of transcription and by hairpins, respectively. The structure of phage λMQ21ΔNB carrying the p_1–*'infC–rpmI'–'lacZ* fusion (a prime on one side of a gene means that it is interrupted on that side) is shown underneath. The boundaries of the region of the operon cloned in front of the *rpmI'–'lacZ* fusion are indicated by two arrows pointing downwards. Mutations are numbered as in Table I. Mutations showing strong (>80% decrease of wild-type repression in Table I) or moderate (decrease comprised between 50 and 80% of wild-type repression in Table I) effects are within black or shaded squares. The ΔApaLI deletion is represented by an open rectangle within *infC*. *iris* is an acronym for *infC-rpmI* intergenic spacer.

(t_1) located at the end of *infC*, and a second terminator (t_2) located at the end of the operon. The expression from this operon is regulated by two control loops acting at the translational level (Lesage *et al.*, 1990). First, IF3 represses the expression of its own gene (Butler *et al.*, 1986). Second, L20 represses the expression of *rpmI* and that of its own gene, *rplT*, via translational coupling with *rpmI* (Lesage *et al.*, 1992). IF3 recognizes the AUU start codon of *infC* (Butler *et al.*, 1987; Sacerdot *et al.*, 1996), whereas the L20 operator extends over ~430 nucleotides, from within the distal half of *infC* to the beginning of *rpmI* (Lesage *et al.*, 1992).

Here we investigate how *rpmI* expression is regulated by L20 over such a distance. We use a combination of *in vivo* and *in vitro* approaches to show that regulation at a distance involves a long-range interaction between the loop of a hairpin structure within *infC* and a region within the *rpmI* translation initiation site, 280 nucleotides downstream.

Results

The experimental system

We have shown previously that L20 represses *rpmI* expression by acting on sequences located between the p_1 promoter within *infC* and the beginning of the *rpmI* coding region (Lesage *et al.*, 1992). Thus, although transcripts from p_1 contribute only marginally to the expression of *rpmI* and *rplT*, they contain all the regulatory regions essential for control by L20. For this reason, we chose to look at the regulation of *rpmI* expressed from the p_1 promoter. Accordingly, genetic screens were performed with a p_1–*'infC–rpmI–rplT'–'lacZα* translational fusion (a prime on one side of a gene means that it is interrupted on that side) cloned in bacteriophage M13 as explained below, and physiological studies of *rpmI* regulation were done exclusively with p_1-*'infC–rpmI'–'lacZ* translational fusions cloned in bacteriophage λ. The control of *rpmI* expression was measured by looking at repression of the *rpmI'–'lacZ* fusions under conditions of L20 excess provided by plasmid pBL6. The wild-type *rpmI'–'lacZ* fusion synthesizes 110 Miller units of β-galactosidase in the presence of the pBR322 control plasmid and 1.7 units in the presence of pBL6 (Table I). The effect of L20 on the *rpmI'–'lacZ* fusions is given as 'repression', which is the ratio of the expression in the presence of pBR322 to that in the presence of pBL6, i.e. 62-fold (Table I). If the expression in the presence of pBL6 is <1.5 units of β-galactosidase, it is possible that this represents 'background' initiation from sites other than the true *rpmI* initiation codon and that repression values are underestimated.

Two distant sites in the mRNA are responsible for the control of rpmI expression

To find the *cis*-acting sites involved in the control of *rpmI* expression by L20, we screened for potential operator-constitutive mutants that derepress the expression of *rpmI*. We screened for operator mutations using a p_1–*'infC–rpmI–rplT'–'lacZα* fusion cloned in M13. This permits easy screening and sequencing of regulatory mutations (see Materials and methods). This screen gave mutations that affected either the coupling between *rpmI* and *rplT* or the control of *rpmI* expression. Mutations in the region responsible for the translational coupling between *rpmI* and *rplT* had been characterized previously (Lesage *et al.*, 1992). In the present work, we focused on the mutations in the translational operator of *rpmI*. Operators containing single mutations were introduced directly upstream of the *rpmI'–'lacZα* fusion on phage M13. In a certain number of cases, a single M13 phage carried several mutations in the translational operator of *rpmI*. To be able to correlate the increase of expression with a change at a specific genetic locus, multiple mutations were separated by digestion of the replicative form of the M13 DNA with appropriate restriction enzymes and substituted with wild-type sequences in order to obtain operators containing single mutations upstream of the *rpmI'–'lacZα* fusion on phage M13. Because all the mutations in a multiple mutant are not necessarily associated with an increase of expression, we expected that some of the mutations, once recloned, would be silent, i.e. have no effect on control. Eighteen out of 21 *rpmI'–'lacZα* mutations we analysed were obtained using this strategy (Table I). Three mutations (infCUAG244, infCUAG280 and AflIIb) were introduced by site-directed mutagenesis. All these mutations were cloned from M13 into bacteriophage λ upstream of an *rpmI'–'lacZ* fusion. The effect of the mutations was measured as explained above. The data (Table I) show that mutations that strongly or moderately affect the control of *rpmI* expression are located in two regions of the translational operator (in black and shaded squares in Figure 1). The first region is located from positions 316

Table I. Effect of the mutations and ΔApaLI deletion on the repression of *rpmI'–'lacZ* fusions (a prime on one side of a gene means that it is interrupted on that side)

Mutation No.	*rpmI'–'lacZ* fusion	Position of mutation	β-Galactosidase (in Miller units/A_{650} of bacteria)			Repression pBR322/pBL6
			IBPC5311	IBPC5311 pBR322	IBPC5311 pBL6	
	λMQ21ΔNB (wild-type)		110 ± 7.9	110 ± 3.5	1.8 ± 0.1	62.5
1	λMQ21–(infCUAG244)ΔNB[c]	infC244 (A→U)	76.1 ± 4.8	92.6 ± 1.9	1.6 ± 0.1	56.8
2	λMQ21–100ΔNB[b]	infC273 (C→U)	96.5 ± 4.1	95.1 ± 1.9	1.9 ± 0.1	51.1
3	λMQ21–(infCUAG280)ΔNB[c]	infC280 (A→U)	92 ± 11.5	109 ± 2.5	1.7 ± 0.1	64.1
4	λMQ21–44ΔNB[b]	infC305 (C→U)	94.9 ± 4.2	111 ± 8.7	1.4 ± 0.2	77.1
5**	λMQ21–146ΔNB[a]	infC316 (G→A)	167 ± 14.4	214 ± 9.4	32.6 ± 1.0	6.5
6**	λMQ21–93ΔNB[a]	infC325 (G→A)	165 ± 5.9	156 ± 19.3	14.9 ± 0.5	10.5
7**	λMQ21–156ΔNB[a]	infC331 (C→U)	431 ± 25.5	503 ± 20.0	54.6 ± 1.3	9.2
8*	λMQ21–111ΔNB[a]	infC336 (C→U)	359 ± 15.4	341 ± 37.1	14.4 ± 0.7	23.7
9*	λMQ21–5ΔNB[b]	infC346 (C→U)	108 ± 19.5	152 ± 4.7	5.7 ± 0.1	26.7
10	λMQ21–75ΔNB[b]	infC355 (G→A)	99.2 ± 2.4	89.9 ± 1.6	1.2 ± 0.1	76.2
11	λMQ21–82ΔNB[a]	infC360 (G→A)	105 ± 7.1	135 ± 4.3	2.1 ± 0.3	64.3
12	λMQ21–9.1.5ΔNB[b]	infC380 (C→U)	123 ± 2.5	112 ± 13.0	1.6 ± 0.1	70.0
13	λMQ21–AflIIbΔNB[c]	infC438 (U→G)	116 ± 8.6	103 ± 11.4	1.1 ± 0.1	96.3
14	λMQ21–173ΔNB[b]	infC516 (G→A)	99.8 ± 12.7	95.1 ± 3.8	1.4 ± 0.2	69.9
15	λMQ21–10ΔNB[b]	infC526 (C→U)	105 ± 11.8	88.0 ± 19.3	1.8 ± 0.2	48.3
16	λMQ21–M5.3ΔNB[a]	infC532 (A→G)	151 ± 24.2	151 ± 16.9	3.1 ± 0.4	48.7
17	λMQ21–3ΔNB[b]	iris18 (G→A)	205 ± 14.3	234 ± 10.0	2.0 ± 0.3	117.0
18*	λMQ21–84ΔNB[a]	iris78 (C→U)	289 ± 12.4	224 ± 9.1	7.2 ± 0.5	31.1
19*	λMQ21–12ΔNB[a]	iris79 (G→A)	291 ± 5.9	271 ± 21.2	13.1 ± 1.8	20.7
20*	λMQ21–92ΔNB[a]	iris82 (G→A)	264 ± 30.6	173 ± 9.4	5.8 ± 0.3	29.8
21	λMQ21–162ΔNB[b]	rpmI18 (C→U)	59.4 ± 8.6	53.4 ± 6.5	0.9 ± 0.1	56.8
		Deletion boundaries				
	λMQ21–ΔApaLIΔNB[c]	ΔinfC363–infC517	88.0 ± 2.8	93.0 ± 2.2	1.7 ± 0.15	54.7

Mutations are numbered according to their positions in the 5′ to 3′ orientation. Numbers followed by a double asterisk indicate that the corresponding mutations produce a >80% decrease of *rpmI* repression compared with the wild-type repression. Numbers followed by a single asterisk indicate that the corresponding mutations result in a decrease of repression comprised between 50 and 80%. Superscripts have the following significance: [a]single mutations obtained by phenotypical screen and directly cloned in front of the *rpmI'–'lacZ* fusion; [b]mutations obtained by phenotypical screen but separated from additional mutations by recombination with wild-type sequences before cloning in front of the fusion; [c]single mutations obtained by site-directed mutagenesis and then cloned in front of the fusion. iris is an acronym for *infC-rpmI* intergenic spacer.

to 346 of *infC*, in the middle of the gene, and the second from nucleotides 78 to 82 of *iris* (*iris* is an acronym for *infC–rpmI* intergenic spacer), that is immediately upstream of *rpmI* (Figure 1). Nucleotides at positions infC346 and iris78 are separated by 274 nucleotides. Mutations that are located between these two regions had little or no effect on the control of *rpmI* expression (Table I and Figure 1). This seems to indicate that this intervening region is not involved in control. This hypothesis was confirmed by the fact that the ΔApaLI deletion, encompassing nucleotides from positions infC363 to infC517, did not affect the control (Table I and Figure 1). Thus, our data indicate that the regions required for the control of *rpmI* expression are located at two distant sites in the translational operator, the first site being located within *infC* and the second upstream of the *rpmI* SD sequence, >270 nucleotides downstream.

Secondary structure probing of the rpmI translational operator reveals the potential existence of a long-range RNA–RNA pseudoknot

The structure of the region between position 214 of *infC*, located just downstream of promoter p_1, and position 10 of *rpmI* (for each gene, position 1 is the first nucleotide of the initiation codon) was investigated using dimethyl sulphate (DMS), 1-cyclohexyl-3-[β-morpholinyl-(4)-ethyl]-carbodiimide methyl-*p*-toluene sulphonate (CMCT) and kethoxal as chemical probes (Peattie, 1979; Inoue and Cech, 1985; Moazed *et al.*, 1986). The secondary structure model of the *rpmI* translational operator proposed in Figure 2 is an adjusted version of that predicted by the RNA MFold program (Zuker, 1989) using the constraints imposed by our biochemical data (see Materials and methods). We have divided the operator into four arbitrary domains. The first domain contains stem–loop structures 1–5. The second domain contains stem–loop structures 6–10, with stem 6 bringing together stem–loop structures 7–10. This second domain is probably not required for the control of *rpmI* expression since point mutations therein and the ΔApaLI deletion, which results in the loss of stem–loop structures 7–10, have almost no effect on control (Table I and Figure 1). The third domain is constituted by the single stem–loop structure 11 which contains the ρ-independent transcriptional terminator t_1 in addition to the translational stop codon UAA of *infC* (positions infC541–infC543). Finally, the fourth domain contains the RBS (ribosomal binding site) of *rpmI* with its SD sequence (UGGA) located from iris83 to iris86. Three supposedly single-stranded sites contain nucleotides exhibiting an unexpected resistance to chemical modification. One such site (sequence UUAAUC from positions infC434 to infC439) is located in the loop of stem–loop structure 9. However, this site is apparently not necessary for the control of *rpmI* expression since it is located in the second domain of the operator which can be deleted without any effect on *rpmI* expression (ΔApaLI deletion

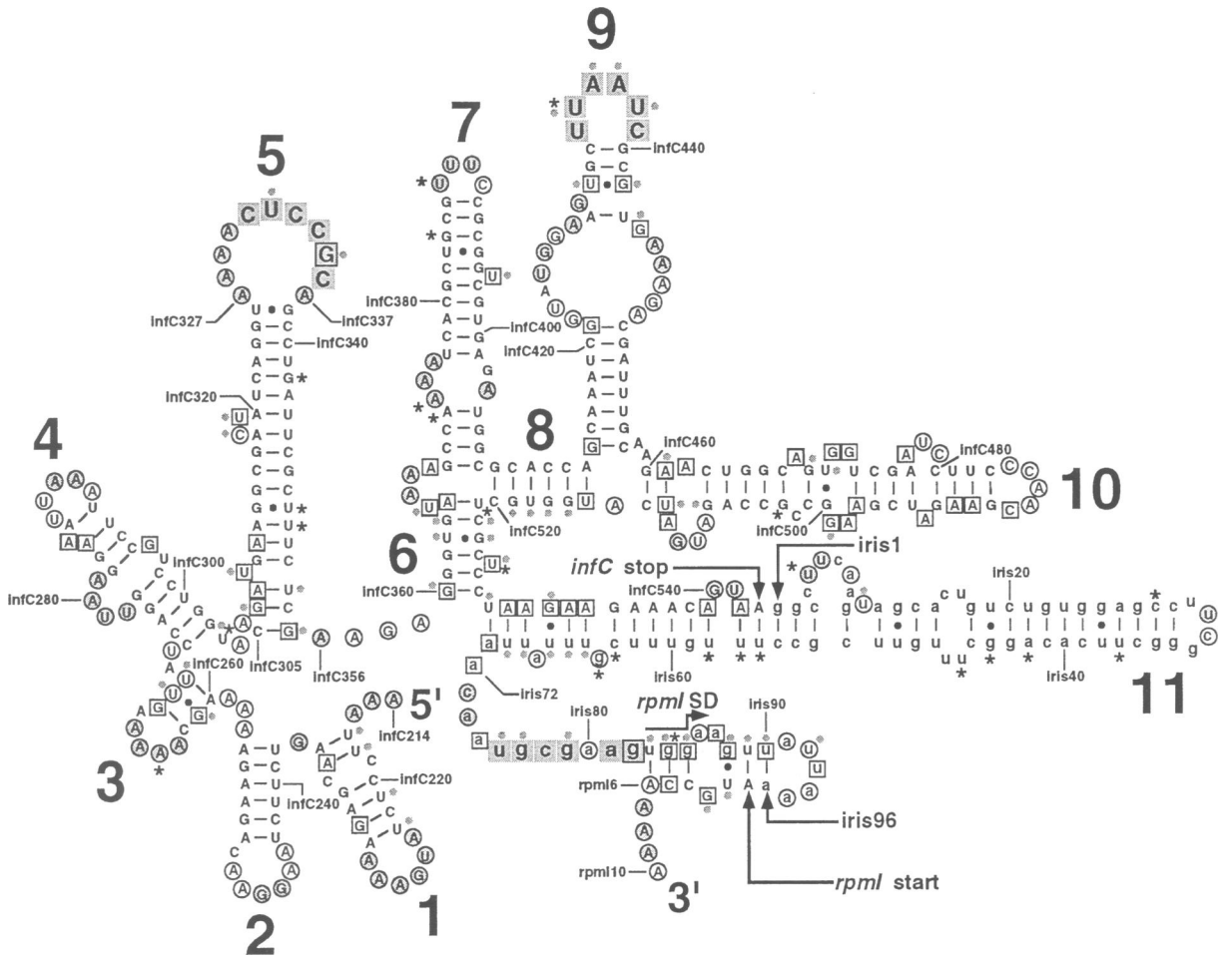

Fig. 2. Secondary structure model of the *rpmI* translational operator. The positions shown are numbered according to the first nucleotide of the corresponding gene (the first nucleotide of the initiation codon is taken as position 1) or *iris*. The *infC* termination and *rpmI* initiation codons are denoted by *infC* stop and *rpmI* start, respectively. The 5′ and 3′ boundaries of *iris* are indicated by arrows pointing to positions iris1 and iris96, respectively. The Shine–Dalgarno sequence of *rpmI* is shown by an arrow indicating the direction of translation and is denoted by *rpmI* SD. *infC* and *rpmI* nucleotides are in upper case and *iris* nucleotides in lower case letters. Positions infC305, infC327, infC337, infC356, iris72, iris90 and rpmI6, which are relevant to Figure 3, are also shown. Stem and stem–loop structures are indicated by numbers. Shaded circles, open circles and squares denote strong, moderate and weak/marginal chemical modifications, respectively. Nucleotides that display unexpected resistance to chemical modification are within shaded squares. Shaded dots indicate enhanced modifications in 'semi-native' conditions. Non-canonical base pairs are represented by black dots. Reverse transcriptase stops and/or pauses are indicated by asterisks.

in Table I and Figure 1). The two other sites are of particular interest. The first (sequence CUCCGC from positions infC331 to infC336) is located in the loop of stem–loop structure 5 and is supposed to be single-stranded in our model. All six of these residues are unreactive to the chemical probes in 'native' conditions, except G(infC335) which exhibits marginal reactivity with kethoxal (Figure 3). In 'semi-native' conditions, U(infC332) becomes reactive with CMCT, and G(infC335) displays enhanced reactivity with kethoxal and is marginally reactive with CMCT (Figure 3). The absence of reactivity of the overall site suggests that this region is base-paired in the secondary structure. The second site of interest (sequence UGCGAAG from iris76 to iris82) is located just upstream of the *rpmI* SD sequence and is also supposed to be single-stranded in our model. However, five out of the seven residues are unreactive to the probes in 'native' conditions (Figure 3). Only A(iris80) and G(iris82) are somewhat reactive. In 'semi-native' conditions, U (iris76) and G residues (iris77 and iris79) become weakly reactive with CMCT and kethoxal, respectively (Figure 3). Taken together, our data suggest that the two sites are not single-stranded in the secondary structure. Since the two sites are partially complementary, we propose they base-pair to form stem S2 in the model shown in the lower part of Figure 3. S2 results from base pairing between nucleotides from positions infC331 to infC337 and nucleotides from positions iris76 to iris82, except at C(infC333) and A(iris80) where a mismatch occurs. This long-range interaction (>280 residues separate the two sites, see Figure 2) results in the formation of a pseudoknot structure. In this pseudoknot structure, the resulting S2 is connected to stem S1 by loops L1 and L2 (Figure 3). Schematized S1 is the stem of hairpin 5 in the secondary structure model shown in Figure 2. L1 contains a single-stranded stretch of four A nucleotides from positions infC327 to infC330. L2 encompasses nucleotides from positions infC356 to iris75 (see Figure 2) and contains stem and stem–loop structures 6–11, amongst which structures 7–10 are not required for the control of *rpmI* expression. Stem S3, which contains the SD sequence and the initiation codon of *rpmI*, is postulated to stack on

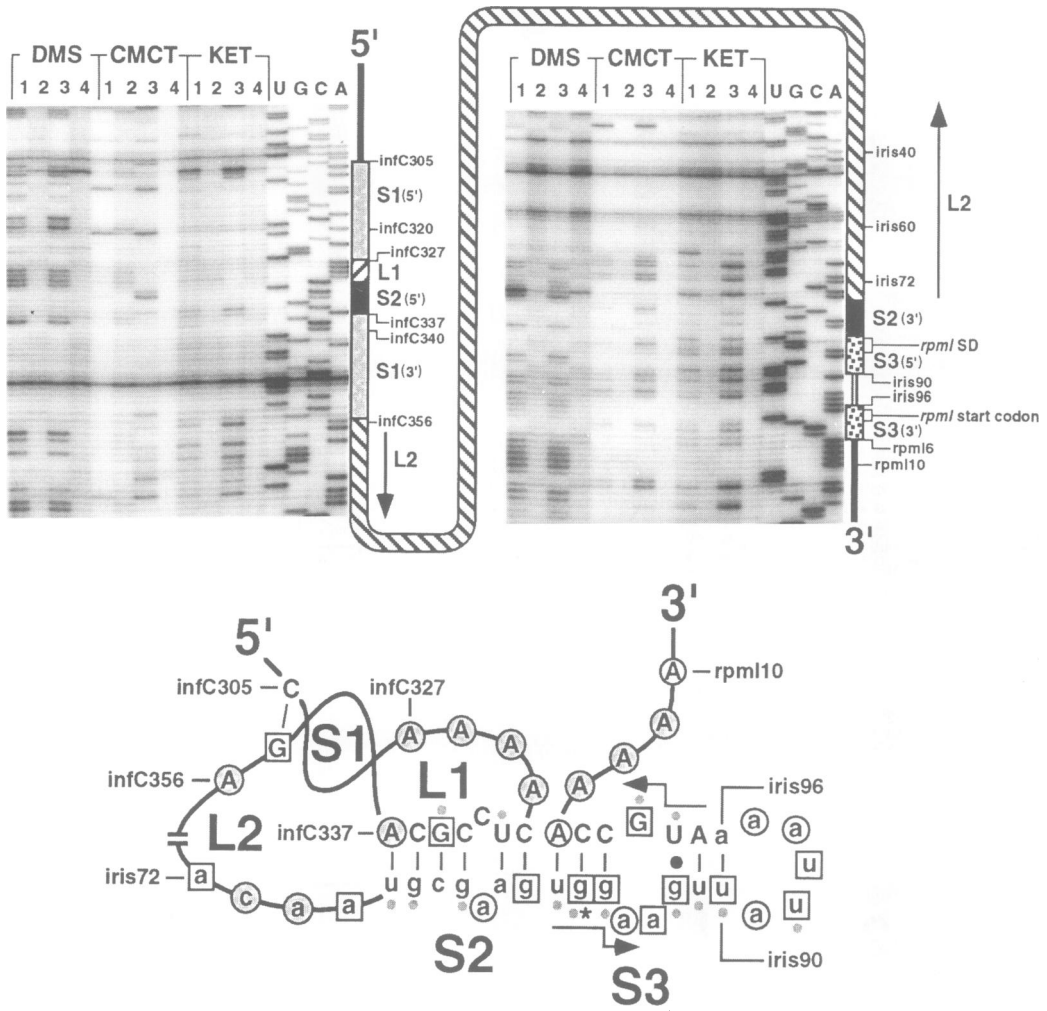

Fig. 3. Primer extension analysis identifying sites of chemical modifications in the proposed pseudoknot. DMS modifies unpaired adenine at N-1 and, more slowly, cytosine at N-3. CMCT modifies unpaired uracil at N-3 and, more slowly, guanine at N-1. Kethoxal (KET) modifies unpaired guanine at N-1 and N-2. Autoradiographs showing primer extension analysis using oligonucleotides DEL2 and rpmI84 are shown in the left and right panels, respectively. Lanes 1 and 3: modifications were performed in 'native' and 'semi-native' conditions, respectively. Lanes 2 and 4 are the corresponding controls, carried out without modifying reagents. Lanes U, G, C and A represent dideoxy sequencing lanes, generated by primer extension of the untreated 'long transcript'. The drawing extending from the right side of the left panel to the right side of the right panel sketches the mRNA. The blocks represent the loops and stems with the same nomenclature as in the bottom figure. The 5′ and 3′ strands of the stems are denoted by 5′ and 3′ between parentheses, respectively. The Shine–Dalgarno sequence and the start codon of *rpmI* are shown by brackets and are indicated by *rpmI* SD and *rpmI* start codon, respectively. The other indicated positions are as in Figure 2. In the drawing at the bottom of the figure, nucleotides of *infC* and *rpmI* are in upper case whereas those of *iris* are in lower case. The Shine–Dalgarno sequence and the initiation codon of *rpmI* are shown by arrows indicating the direction of translation. Intensities and enhancements of modifications, reverse transcriptase stops and/or pauses and non-canonical base pairs are represented as in Figure 2. Stem S1 is the stem of stem–loop structure 5 (see Figure 2). Stem S2 results from base pairing of nucleotides from positions infC331 to infC337 with nucleotides from positions iris76 to iris82, located upstream of the *rpmI* SD (see Figure 2). The loops connecting S1 and S2 are indicated by L1 and L2. L1 contains four A residues from positions infC327 to infC330 (see Figure 2). L2 consists of nucleotides from infC356 to iris75 (see Figure 2). Stem S3 is formed by base pairing of nucleotides from iris83 to iris90 with nucleotides from iris96 to rpmI6 (see Figure 2).

S2, thus extending even further the quasi-continuous double helix generated by the pseudoknot (Figure 3). An intriguing result is the reactivity of A(infC337) towards DMS and the lack of reactivity of U(iris76) towards CMCT (Figure 3). This result is clearly in disagreement with a Watson–Crick base pairing since N-1 of adenine and N-3 of uracil should be engaged in hydrogen bonding and not be reactive towards these chemicals. However, since our results indicate that N-1 of adenine seems to be accessible whereas N-3 of uracil does not, we cannot exclude that A(infC337) and U(iris 76) base-pair by reverse Hoogsteen interaction as does the A18·U8 interaction found in crystallographic studies of yeast $tRNA^{Phe}$ and $tRNA^{Asp}$ (Quigley *et al.*, 1975; Jack *et al.*, 1976; Westhof *et al.*, 1985). Interestingly, as in the case of the A(infC337)·U(iris76) base pair in S2, this A18·U8 base pair is at the far end of a stem (the D-stem) consisting of Watson–Crick base pairs and allows a tertiary interaction to occur.

The long-range RNA–RNA tertiary interaction in the translational operator is required for control of rpmI expression

In order to provide genetic evidence for the existence of such a long-range RNA–RNA interaction and its involvement in control of *rpmI* expression, a mutational

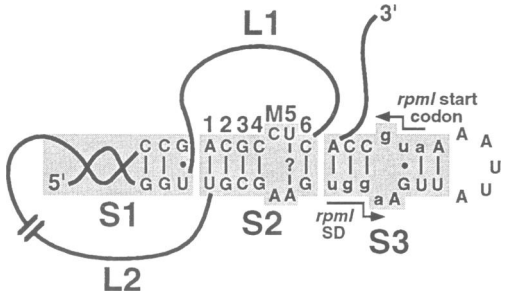

Fig. 4. Model of the proposed pseudoknot. Stems S1, S2 and S3 are in shaded boxes. The stems and the loops are represented as in Figure 3. Base pairs forming stem S2 are indicated by numbers and the C·A mismatch by the letter M. Base pair 5 is also indicated by a question mark which means that it is not substantiated by compensatory base changes. The Shine–Dalgarno sequence and the initiation codon of *rpmI* are in lower case and are denoted by *rpmI* SD and *rpmI* start codon, respectively. The arrows indicate the direction of translation.

analysis was carried out on the pseudoknot key stem, S2. The effect of mutations on *rpmI* expression was investigated *in vivo* by β-galactosidase activity measurements from a *rpmI'–'lacZ* fusion as described above. We analysed the effect of mutations in the two complementary strands of the putative S2 (Figure 4). All of the mutations, but one (mutation 2b* which corresponds to a C to U change at infC336), were introduced by site-directed mutagenesis. Most of these mutations were expected to disrupt the base pairing in the stem and to result in a decreased control of *rpmI* expression. Disruption of one of the three G–C base pairs (numbered 2, 3 and 4 in Figure 4) located in the left side of S2 strongly affect the control (Table II). Mutations introduced in the three other predicted base pairs (numbered 1, 5 and 6 in Figure 4) also affect control, although to a lesser extent. Mutation 1* at position infC337 has a negligible effect on the control. This mutation probably does not cause a base pair disruption because a U·G wobble base pair can form. Since each base was changed to the complementary partner at every position in the predicted S2, the combination of two changes should restore base pairing. With the compensatory mutants (1*/1**), (2a*/2a**), (3*/3**) and (6*/6**), repression of *rpmI* expression was completely (or almost completely) recovered (Table II). The compensatory mutant (2b*/2b**) which converts a G–C to an A–U base pair, displays a partial (~60% of the wild-type repression) recovery of control. The two compensatory mutants (4*/4**) and (5*/5**), which correspond to changes of nucleotides flanking the A(iris80)·C(infC333) mismatch (Figure 4), show only partial restoration of the repression (~50% of the wild-type repression) in the first case and no restoration at all in the second. As a control, a double mutant (2b*/1**) was also constructed by combining two changes of nucleotides that are predicted not to base-pair. As expected, this double mutant does not restore the repression but rather results in an additive effect of the two single mutations on the repression of *rpmI* expression.

In conclusion, apart from the A–U base pair 5 in Figure 4, the loss of regulation caused by every change in either strand of S2 can be compensated by a complementary change in the other strand. This is a good indication that S2 really forms *in vivo* and strongly supports the finding that a pseudoknot structure is required for *rpmI* repression by L20.

Since control is affected by every change that disrupts S2, one would expect that the A to G change in the A(iris80)·C(infC333) mismatch, which eliminates the mismatch and stabilizes S2, would not affect control unless one of these two nucleotides is involved in some specific recognition event. In fact, this mutation (called M* in Table II) causes a decrease in expression and a slight decrease in control. However, because of the low level of expression under repressed conditions, we suspected that the repression is underestimated (see above). Complementary experiments under conditions where L20 is only slightly overproduced show the same 5-fold repression for the wild-type and the A to G (iris80) mutated fusions (data not shown). This indicates that the M* mutation has no effect on control.

Discussion

Features of the pseudoknot formed by the long-range RNA–RNA interaction in the rpmI operator

Our results show that only two sites within the 430 nucleotides of the upstream untranslated region are required for efficient control of *rpmI* expression by ribosomal protein L20. The first is located within *infC* at the top of hairpin structure 5, and the second just upstream of the *rpmI* SD sequence (Figure 2). Mutations in these sites (shown as black and shaded squares in Figure 1) affect the control of *rpmI* expression. Structural probing indicates that nucleotides in the loop of hairpin 5 of the first site pair with nucleotides of the second site. This interaction takes place over ~280 nucleotides to form stem S2 of the pseudoknot shown in Figures 3 and 4. That this structure exists *in vivo* is shown by the fact that single changes that disrupt S2 and abolish control can be compensated by changes that re-establish base pairing.

The pseudoknot structure inferred by our mutational analysis is based upon a stacking interaction between S2, a non-continuous helix with an A·C mismatch, and S1, an 18 bp helix containing several bulged nucleotides (see the stem of hairpin 5 in Figure 2). The existence of such a structure could result in the formation of a quasi-continuous extended double helix resulting from a stacking interaction between S1 and S2 (Figure 4). In this extended double helix, the two helical regions are connected by L1 and L2. This stacking interaction results in the formation of an H-type pseudoknot (Pleij *et al.*, 1985; Pleij and Bosch, 1989) where the two stems are connected by L1 and L2. According to current pseudoknot building models (Pleij *et al.*, 1985; Pleij and Bosch, 1989), the lengths of L1 (four nucleotides) and L2 (>260 nucleotides) make the bridging of the two stems sterically feasible. Pseudoknots with long L2s are well characterized in rRNAs (Gutell, 1993).

Although the interpretation of the biological effects caused by the introduction of changes to an irregular helix, in particular at positions close to mismatches, internal loops or bulges, has to be made with care, the effect on regulation of the different mutations in S2 seems to be quite well correlated with their effect on the free energy of helix formation, as calculated using the nearest-

Table II. Single and double mutations in the proposed pseudoknot and their effect on the repression of *rpmI'–'lacZ* fusions (a prime on one side of a gene means that it is interrupted on that side)

Mutation No.	*rpmI'–'lacZ* fusion	Position of mutation	β-Galactosidase (in Miller units/A_{650} of bacteria)			Repression pBR322/pBL6	$\Delta\Delta G°$ (kcal/mol)
			IBPC5311	IBPC5311 pBR322	IBPC5311 pBL6		
	λMQ21ΔNB (wild-type)		110 ± 8.0	110 ± 3.5	1.8 ± 0.1	62.5	0
1*	λMQ21-337infCΔNB	infC337 (A→G)	162 ± 20.3	156 ± 10.5	2.9 ± 0.1	53.8	+0.3
1**	λMQ21-76SIRΔNB	iris76 (U→C)	246 ± 28.5	187 ± 14.9	5.4 ± 0.1	34.6	+3.1
1*/1**	λMQ21-(337infC/ 76SIR)ΔNB	infC337 (A→G)/ iris76 (U→C)	110 ± 14.6	97.7 ± 5.7	1.5 ± 0.4	66.4	−1.0
2a*	λMQ21-336infCΔNB	infC336 (C→G)	485 ± 6.4	257 ± 15.5	35.4 ± 1.7	7.2	+6.0
2a**	λMQ21-77'SIRΔNB	iris77 (G→C)	392 ± 18.8	409 ± 13.6	30.5 ± 2.4	13.4	+6.0
2a*/2a**	λMQ21-(336infC/ 77'SIR)ΔNB	infC336 (C→G)/ iris77 (G→C)	191 ± 13.8	252 ± 21.5	4.8 ± 0.4	52.5	0
2b*	λMQ21-111ΔNB	infC336 (C→U)	359 ± 15.4	341 ± 37.1	14.4 ± 0.7	23.7	+2.6
2b**	λMQ21-77SIRΔNB	iris77 (G→A)	502 ± 4.5	209 ± 9.2	24.5 ± 1.3	8.5	+6.0
2b*/2b**	λMQ21-(111/77SIR)ΔNB	infC336 (C→U)/ iris77 (G→A)	404 ± 33.3	310 ± 36.2	7.3 ± 0.3	42.5	+2.0
3*	λMQ21-335infCΔNB	infC335 (G→C)	256 ± 12.6	309 ± 5.4	27.2 ± 2.0	11.4	+6.2
3**	λMQ21-78SIRΔNB	iris78 (C→G)	384 ± 14.9	292 ± 27.6	15.3 ± 0.7	19.1	+6.2
3*/3**	λMQ21-(335infC/ 78SIR)ΔNB	infC335 (G→C)/ iris78 (C→G)	104 ± 19.5	122 ± 7.6	2.0 ± 0.2	61.0	−0.4
4*	λMQ21-334'infCΔNB	infC334 (C→G)	321 ± 13.2	233 ± 10.3	12.7 ± 0.9	18.3	+2.9
4**	λMQ21-79SIRΔNB	iris79 (G→C)	321 ± 21.6	284 ± 22.5	12.9 ± 0.7	22.0	+2.9
4*/4**	λMQ21-(334'infC/ 79SIR)ΔNB	infC334 (C→G)/ iris79 (G→C)	121 ± 3.5	116 ± 10.9	3.4 ± 0.2	34.2	−0.9
M*	λMQ21-80'SIRΔNB	iris80 (A→G)	33.4 ± 3.3	65.7 ± 7.7	1.4 ± 0.1	47.9	−6.0
5*	λMQ21-332infCΔNB	infC332 (U→A)	247 ± 4.2	339 ± 11.0	7.1 ± 0.4	47.7	+2.6
5**	λMQ21-81SIRΔNB	iris81 (A→U)	289 ± 11.0	285 ± 23.2	6.3 ± 0.3	45.2	+2.6
5*/5**	λMQ21-(332infC/ 81SIR)ΔNB	infC332 (U→A)/ iris81 (A→U)	276 ± 17.6	270 ± 17.2	11.1 ± 0.3	24.3	−0.1
6*	λMQ21-331infCΔNB	infC331 (C→G)	424 ± 11.0	273 ± 20.5	11.8 ± 1.2	23.1	+4.6
6**	λMQ21-82SIRΔNB	iris82 (G→C)	265 ± 17.6	206 ± 11.5	6.8 ± 1.3	30.3	+4.6
6*/6**	λMQ21-(331infC/ 82SIR)ΔNB	infC331 (C→G)/ iris82 (G→C)	109 ± 8.4	110 ± 5.1	2.1 ± 0.3	52.4	0
2b*	λMQ21-111ΔNB	infC336 (C→U)	359 ± 15.4	341 ± 37.1	14.4 ± 0.7	23.7	+2.6
1**	λMQ21-76SIRΔNB	iris76 (U→C)	246 ± 28.5	187 ± 14.9	5.4 ± 0.1	34.6	−1.0
2b*/1**	λMQ21-(111/76SIR)ΔNB	infC336 (C→U)/ iris76 (U→C)	408 ± 2.8	325 ± 11.9	23.8 ± 0.3	13.7	+4.6

Mutations are numbered according to the numbering of the corresponding base pair in stem S2 as shown in Figure 4. The last column shows the corresponding theoretical changes in the free energy of S2 formation: $\Delta\Delta G°$ = $\Delta G°$(mutant) – $\Delta G°$(wild-type). $\Delta G°$ of wild-type (–20.2 kcal/mol) and mutated S2 were obtained by applying the thermodynamic parameters of Freier *et al.* (1986) to the calculation of the stability at 37°C of a continuous helix in which S2 is extended by three base pairs belonging to S1 on one side and by three base pairs belonging to S3 on the other as shown in Figure 4.

neighbour rules (see $\Delta\Delta G°$ values for S2 mutations relative to wild-type in Table II). The changes at the G–C base pairs 2 and 3 (Figure 4), which have the strongest effect on regulation (Table II), correspond to the least stable structures. For most of the mutations of these two base pairs, control is strongly affected (<30% residual control) and S2 stability is relatively low (their associated $\Delta\Delta G°$ is equal to or higher than +6 kcal/mol). The only exception is the G–C to G·U change at base pair 2 (mutation 2b*) for which control is only affected to a limited extent (~40% residual control) and the stability of S2 remains reasonably high ($\Delta\Delta G°$ = +2.6 kcal/mol). The correlation between control and S2 stability also holds in the case of changes in the G–C base pair 4, since a limited effect on control (~30–35% residual control) is associated with a reasonably high stability ($\Delta\Delta G°$ = +2.9 kcal/mol) which is comparable (see Table II) with that of mutation 1** at base pair 1 and mutation 2b* at base pair 2. The changes at base pairs 1, 5 and 6, which only slightly affect control, have a limited effect on S2 stability since their associated $\Delta\Delta G°$s are between those of mutations which strongly affect S2 stability and that of the wild-type. The only clear exception where the effect on regulation of a mutation cannot be correlated with an effect on S2 stability is the A·C mismatch to G–C change (mutation M* in Table II). This mutation causes a strong stabilization of S2 ($\Delta\Delta G°$ = –6.0 kcal/mol) without increase of control. This result could simply mean that S2 is stable enough in the wild-type operator to fold it in a correct conformation. Our results clearly indicate that the nucleotides of S2 are involved in regulation only in an indirect way. Rather than making essential contacts with other regions of the RNA or with L20, the S2 nucleotides seem to be involved in the folding of the operator so that other nucleotides of the RNA are positioned correctly to perform an essential role. This is indicated by the fact that any mutation in S2 affecting control can be compensated by a change in the opposite strand, suggesting no apparent sequence specificity. Compensation does not work in the case of the A–U base pair 5 (Figure 4 and Table II) which could mean that one or other of these nucleotides has a specific role. However, this seems unlikely since the effects of

single changes in this base pair are weak. The A of the A·C mismatch in S2 is also not specifically involved in regulation since a conversion to a G–C base pair does not abolish control.

Working model for rpmI regulation

Although we suspect that L20 binds to the *rpmI* operator, we have no direct experimental evidence for such an interaction. In other translational operators, such as those of the *rpsO* ribosomal protein gene and the α-operon, the pseudoknots are involved directly in binding their respective repressors (Tang and Draper, 1989; Philippe *et al.*, 1990). A subset of mutations in these pseudoknots consistently abolishes control. None of the mutations we have isolated in the *rpmI* pseudoknot have such a drastic effect. The mutation that shows the strongest effect decreases the repression value from 62.5 to 6.5 and changes a nucleotide located in stem S1 (mutation No. 5 in Table I). Therefore, we conclude that none of the mutations we isolated in the pseudoknot changes a nucleotide performing a task essential for control. This suggests that L20 might bind to other sites in the operator in addition to nucleotides located in sequences forming the pseudoknot. In the original model of Nomura and collaborators, the control ribosomal proteins were proposed to recognize similar sequences on their mRNA and rRNA binding sites. In this respect, an AAGAGGG sequence is found both in the *rpmI* translational operator (positions infC356 to infC362 in Figure 2) and in a region of 23S rRNA that is protected from RNase T1 attack by L20 (positions 972–978) (C.Branlant, unpublished results). The three consecutive G residues at the end of the AAGAGGG sequence that base-pair to form stem 6 (Figure 2) are also base-paired in the L20 binding site of 23S rRNA. We have no indication that the single-stranded AAGA sequence (positions infC356 to infC359) is involved in control since no mutation was found in any of these nucleotides. However, we can exclude that the double-stranded GGG sequence is involved in control since the 82 mutation at G(infC360), which disrupts the base pair at the bottom of stem 6, has no effect on the repression level (Table I). Also, the ΔApaLI deletion, which has little effect on the control (Table I), results in the disruption of the base pair at the top of stem 6. However, the paired GGG sequence may be required for bringing together stems and stem–loop structures 7–10. Therefore, L20 may bind to the AAGA sequence, to nucleotides located in the sequences forming the pseudoknot or even elsewhere in the *rpmI* leader mRNA. It is possible that the pseudoknot stabilizes a conformation of the mRNA that allows the AAGA sequence to be close to the RBS of *rpmI*, thus reconstituting an L20 binding site with sequences from scattered regions of the operator.

So far, we have no idea of the mechanism by which L20 regulates *rpmI* expression. The repressor could, once bound to the operator, either inhibit the binding of the ribosome (Winter *et al.*, 1987; Hartz *et al.*, 1989; Moine *et al.*, 1990) or trap it in an inactive initiation complex (Philippe *et al.*, 1993; Spedding *et al.*, 1993), as with other genes regulated at the translational level. Another possibility is that L20 is essential for stacking S1, S2 and S3 on top of each other. The formation of the pseudoknot, as underlined in Results and illustrated in Figure 4, could stabilize stem S3 by coaxial stacking. Since S3 sequesters both the *rpmI* SD sequence and translational initiation codon (Figures 3 and 4), it is possible that the formation of the pseudoknot inhibits ribosome binding to the *rpmI* RBS through S3 stabilization. This role for S3 in regulation is suggested by the fact that a change in the SD sequence of *rpmI* that destabilizes S3 also affects regulation (data not shown). In this respect, it is interesting to note that an inhibitory secondary structure sequestering both the SD sequence and translational initiation codon of the MS2 phage replicase gene was proposed to be stabilized by coaxial stacking of a long-range secondary structure (van Himbergen *et al.*, 1993). Thus, once the S1–S2–S3 stacking is formed, translation initiation is likely to be inhibited. The finding that mutation M*, which strongly stabilizes S2 (Table II), decreases *rpmI* expression ~3- to 4-fold (33.4 units of β-galactosidase versus 110 units for the wild-type lysogen) without affecting control (repression factor of 47.9 versus 62.5 for the corresponding pBL6-carrying lysogens) seems to indicate that stabilization of S2 increases the inhibition of ribosome binding to the *rpmI* RBS. Thus, since the S1–S2 stacking, and, possibly, the S1–S2–S3 stacking, is inhibitory by itself, the picture that emerges is that L20 does not seem to inhibit ribosome binding directly or to trap it in an inactive initiation complex but rather helps the operator to adopt the proper conformation for promoting regulation of *rpmI* expression. In this case, L20 could act either as a classical translational repressor by stabilizing a structure that inhibits ribosomal binding or, more appealingly, as an RNA chaperone by facilitating the formation of the S1–S2–S3 stacking which, once formed, would be sufficiently stable, even in the absence of L20, to inhibit ribosome binding. Interestingly, recent *in vitro* experiments indicate that some ribosomal proteins (S12 in particular) seem to be able to behave as RNA chaperones, i.e. facilitate RNA folding to an active conformation (Coetzee *et al.*, 1994). Strangely enough, the same experiments indicate that L35, the product of the *rpmI* gene, is endowed with a weak RNA chaperone activity. We presently are testing the effect of L35 and L20 on the folding of the *rpmI* translational operator to find out whether L20, in the presence or absence of L35, behaves as a classical repressor or as an RNA chaperone.

It is important to mention that our experiments were performed on translational fusions expressed from the p_1 promoter, i.e. translation of *infC* mRNA cannot occur since the fusions are lacking the RBS and the whole N-terminal region of *infC* (Figure 1). Recent experiments using a *rpmI'–'lacZ* fusion containing the $p_{0'}$ promoter, and thus also containing all the sequences necessary for translation of *infC* mRNA, have shown that the regions involved in control on the $p_{0'}$ transcript are globally the same as those important for control on the p_1 transcript, with the difference that the efficiency of repression is much lower in the former case (10-fold versus 62-fold repression under the same conditions, respectively). The lower efficiency of repression on the $p_{0'}$ transcript was shown to be due to the transient melting of the pseudoknot by translating ribosomes or any other structure in the *infC* mRNA essential for an interaction with L20 (C.Chiaruttini, unpublished results). Since ribosomes translating the *infC* mRNA decrease the efficiency of repression, any increase in the frequency of translation initiation of the *infC* mRNA

causes an increase of *rpmI* expression, i.e. the expression of *rpmI* and *rplT* is coupled to that of *infC* (C.Chiaruttini, in preparation).

In conclusion, the pseudoknot described here seems to be involved in L20-mediated translational regulation of *rpmI* on the transcript starting at p_1 and in the coupling between *infC* and *rpmI* on the transcript starting at $p_{0'}$. In addition, this is, to our knowledge, the first report of a pseudoknot structure built up by sites separated by ~280 nucleotides on a prokaryotic mRNA. Pseudoknots involving long-range base pairings have already been well characterized in rRNAs (Gutell, 1993). Although the precise role of the pseudoknot remains to be investigated, we believe that, apart from its role in stabilizing a secondary structure which inhibits translation, it also permits the orientation of specific binding motifs to facilitate interaction with other regions of the operator, with the repressor or other elements that may be involved in the control.

Materials and methods

Strains and standard techniques

The *E.coli* strains, bacteriophages and plasmids used in this work are listed in Table III. For reasons of space, all the M13mp18MQ21ΔNB derivatives which were used for subsequent cloning in λ are not listed. General bacteriological techniques were as described (Miller, 1972; Davis *et al.*, 1980). General cloning techniques were as in Sambrook *et al.* (1989).

Mutagenesis

Site-directed mutagenesis using oligonucleotides was performed as previously described (Kunkel *et al.*, 1987). M13mp18MQ21ΔNB, the M13mp18 derivative carrying the *rpmI'–'lacZ*α fusion expressed from the p_1 promoter (Lesage *et al.*, 1992), was employed as template, except in the cases of mutations infCUAG244 and infCUAG280 for which M13mp18MIRΔNB, the M13mp18 derivative carrying the *rpmI'–'lacZ*α fusion starting upstream of the $p_{0'}$ promoter (Lesage *et al.*, 1992), was used. Mutagenic oligonucleotides were synthesized by OligoExpress (Montreuil, France); sequences of these oligonucleotides are available upon request. The presence of all the mutations was checked using the dideoxy chain termination method (Sanger *et al.*, 1977).

In vitro mutagenesis on the replicative form of M13mp18MQ2114.2 and M13mp18MQ21PL09 with hydroxylamine was performed as already described (Lesage *et al.*, 1992). *In vivo* mutagenesis on M13mp18-MQ2114.2 with JM101*mutD5* was carried out as previously described (Brunel *et al.*, 1992).

Screen for increased expression with M13

We used derivatives of M13mp18MQ21 carrying a p_1–*'infC–rpmI–rplT'–'lacZ*α translational fusion for the genetic screening experiments. In this construct, *rplT* is fused in-phase to the α fragment of *lacZ* and is expressed from the p_1 promoter (Lesage *et al.*, 1992). Consequently, with a strain providing the Ω complementing fragment of *lacZ*, such a phage makes blue plaques on an X-gal indicator plate. Since the expression of *rplT* is translationally coupled to that of *rpmI* (Lesage *et al.*, 1992), any mutation that affects *rpmI* expression can be detected with the p_1–*'infC–rpmI–rplT'–'lacZ*α fusion. Since regulatory mutations that act *in cis* are expected to increase expression, we used M13mp18MQ21 derivatives making light blue plaques as starting material. One of these derivatives, M13mp18MQ2114.2, was constructed by mutating the *rpmI* SD sequence (UGGA to UAAA) which decreased *rpmI* expression ~10-fold and, due to the translational coupling, also decreased that of the *rplT'–'lacZ*α hybrid gene (Lesage *et al.*, 1992). We also used a second derivative, phage M13mp18MQ21PL09, which was constructed by mutating the *rplT* SD sequence (AGGAG to AGUAG) and also expressed the *rplT'–'lacZ*α hybrid gene at a low level (Lesage, 1991). After mutagenesis of the replicative form of these phages *in vitro* with hydroxylamine or *in vivo* in a *mutD5* strain, we screened for phages that gave dark blue plaques, i.e. phages that displayed increased expression of the *rplT'–'lacZ*α hybrid gene. Mutations were sequenced using the termination chain method as described above.

Construction of M13mp18 derivatives carrying rpmI'–'lacZα fusions

As explained in Results, the screen with M13mp18MQ2114.2 and M13mp18MQ21PL09 gave mutations in the *rpmI* operator and in the structure responsible for the translational coupling between *rpmI* and *rplT*. Single mutations isolated using the screen were recloned upstream of the *rpmI'–'lacZ*α fusion in M13mp18MQ21ΔNB. Mutations located downstream of the *Bbs*I site within *infC* and upstream of the *Stu*I site at the end of *infC* were cloned by replacing the wild-type 0.31 kb *Bbs*I–*Stu*I fragment of M13mp18MQ21ΔNB with the corresponding mutated fragment. Mutations located downstream of the *Stu*I site were cloned in M13mp18MQ21ΔNB by a two-step procedure. Since they were all derived from M13mp18MQ2114.2, the wild-type *rpmI* SD sequence was first restored by site-directed mutagenesis with the appropriate oligonucleotide. In the second step, the mutations were cloned into M13mp18MQ21ΔNB by replacement of the wild-type 0.55 kb *Bbs*I–*Nco*I or 0.42 kb *Bbs*I–*Spl*I fragment with the corresponding mutated fragment. Mutations infCUAG244 and infCUAG280, which were introduced into M13mp18MIRΔNB by site-directed mutagenesis, were cloned by replacing the wild-type 0.31 kb *Bbs*I–*Stu*I fragment of M13mp18MQ21ΔNB with the corresponding mutated fragment from M13mp18MIRΔNB. M13mp18MQ21ΔNB derivatives carrying double compensatory mutations were constructed by replacing the wild-type 0.31 kb *Bbs*I–*Stu*I fragment of the derivatives carrying a single mutation downstream of the *Stu*I site with the same fragment containing the appropriate mutation. M13mp18MQ21-ΔApaLIΔNB was constructed by deleting DNA between two *Apa*LI sites (*Apa*LIa and *Apa*LIb) at positions 362 and 517 in *infC*. The two sites were introduced independently into M13mp18MQ21ΔNB by site-directed mutagenesis. The ΔApaLI deletion was made by ligating the 0.3 kb *Apa*LI–*Hin*dIII fragment derived from M13mp18MQ21-ApaLIbΔNB to the 7.4 kb *Apa*LIa–*Hin*dIII fragment derived from M13mp18MQ21-ApaLIaΔNB. All clonings were checked by dideoxy sequencing from the *Hin*dIII site at the *rpmI'–'lacZ* boundary to the *Eco*RI site upstream of the p_1 promoter.

Lambda bacteriophage constructions and β-galactosidase measurements

All the mutations isolated were first transferred to M13mp18MQ21ΔNB, which carries the p_1–*'infC–rpmI'–'lacZ*α fusion, and the fusions were subsequently cloned into the b2 region of bacteriophage λ as previously described (Lesage *et al.*, 1992) to give derivatives of λMQ21ΔNB (Figure 1). In these constructs, *lacZ* was fused in-phase with the first 157 nucleotides of *rpmI*. The fusions subsequently were integrated into the *E.coli* chromosome as single copies at the *att*λ site. Subsequent lysogenization of the *E.coli* strain IBPC5311 and monolysogens screening were performed as described (Springer *et al.*, 1985, 1986). The control of *rpmI* expression was measured by comparing β-galactosidase activities of the monolysogens in the presence of plasmid pBL6 and in the presence of plasmid pBR322 as a control. Plasmid pBL6, which derives from pBR322, expresses *rplT* from the *lac* promoter (Lesage *et al.*, 1992). Since all the upstream regulatory sequences and *rpmI* are absent in this plasmid, *rplT* expression is not under feedback control. Also, it was constructed in such a way that *rplT* translation occurs independently of that of *rpmI*. Monolysogens carrying plasmid pBL6 overexpress L20 by a factor of three with respect to monolysogens carrying pBR322 control plasmid. The presence of equivalent amounts of L20 in all pBL6-carrying monolysogens was checked by immunoblotting (data not shown). Growth conditions of plasmid-carrying monolysogens and β-galactosidase measurements were as previously described (Lesage *et al.*, 1992). β-Galactosidase measurements were carried out at least in duplicate.

Chemical probing experiments

In vitro synthesis of the so-called 'long transcript' with T7 RNA polymerase was carried out as already described (Lesage *et al.*, 1992). This transcript is 43 nucleotides longer than the *in vivo* $p_{0'}$ transcript. The first 33 nucleotides at the 5′ end of the transcript are sequences from the vector and the next 10 nucleotides correspond to the distance between the –10 region of the $p_{0'}$ promoter and its transcription initiation site. The addition of such a sequence to the 5′ end of the natural $p_{0'}$ transcript has no influence on our results since probing experiments performed on the 'short transcript', that initiates upstream of the p_1 promoter (Lesage *et al.*, 1992), gave the same reactivities for the nucleotides shown in Figure 2. The 'long transcript' (0.5 μg equivalent

Table III. Strains of *E.coli*, bacteriophages and plasmids

Name	Genotype	Reference
Escherichia coli		
IBPC5311	*thi-1, argE3, ΔlacX74, galK2?, mtl-1 xyl-5, tsx-29?, rpsL, recA1, pps*	Springer *et al.* (1985)
JM101TR	*F' lacI^q, lacZΔM15, traD36, proAB+, thi, Δ(lac-pro), supE, recA(56?), srl-(300?)::*Tn*10*	Lesage *et al.* (1992)
JM101TR mutD5	*F' lacI^q, lacZΔM15, traD36, proAB+, thi, Δ(lac-pro), supE, recA(56?),mutD5*	Brunel *et al.* (1992)
M13 derivatives		
*rpmI'–'lacZ*α fusions		
M13mp18MQ21ΔNB	*'infC, rpmI'–'lacZ*α	Lesage *et al.* (1992)
M13mp18MIRΔNB	*'thrS, infC, rpmI'–'lacZ*α	Lesage *et al.* (1992)
*rplT'–'lacZ*α fusions		
M13mp18MQ2114.2	*'infC, rpmI14.2, rplT'–'lacZ*α	Lesage *et al.* (1992)
M13mp18MQ21PL09	*'infC, rpmI, rplTPL09'–'lacZ*α	Lesage (1991)
Lambda derivatives		
rpmI'–'lacZ fusions		
λMQ21ΔNB	*cI857, nin5; 'infC, rpmI'–'lacZ*	Lesage *et al.* (1992)
λMQ21-(infCUAG244)ΔNB	*cI857, nin5; 'infCUAG244, rpmI'–'lacZ*	this work
λMQ21-100ΔNB	*cI857, nin5; 'infC100a, rpmI'–'lacZ*	this work
λMQ21-(infCUAG280)ΔNB	*cI857, nin5; 'infCUAG280, rpmI'–'lacZ*	this work
λMQ21-44ΔNB	*cI857, nin5; 'infC44a, rpmI'–'lacZ*	this work
λMQ21-146ΔNB	*cI857, nin5; 'infC146, rpmI'–'lacZ*	this work
λMQ21-93ΔNB	*cI857, nin5; 'infC93, rpmI'–'lacZ*	this work
λMQ21-156ΔNB	*cI857, nin5; 'infC156, rpmI'–'lacZ*	this work
λMQ21-111ΔNB	*cI857, nin5; 'infC111, rpmI'–'lacZ*	this work
λMQ21-5ΔNB	*cI857, nin5; 'infC5a, rpmI'–'lacZ*	this work
λMQ21-75ΔNB	*cI857, nin5; 'infC75a, rpmI'–'lacZ*	this work
λMQ21-82ΔNB	*cI857, nin5; 'infC82, rpmI'–'lacZ*	this work
λMQ21-9.1.5ΔNB	*cI857, nin5; 'infC9.1.5a, rpmI'–'lacZ*	this work
λMQ21-AflIIbΔNB	*cI857, nin5; 'infCAflIIb, rpmI'–'lacZ*	this work
λMQ21-173ΔNB	*cI857, nin5; 'infC173a, rpmI'–'lacZ*	this work
λMQ21-10ΔNB	*cI857, nin5; 'infC10a, rpmI'–'lacZ*	this work
λMQ21-M5.3ΔNB	*cI857, nin5; 'infCM5.3, rpmI'–'lacZ*	this work
λMQ21-3ΔNB	*cI857, nin5; 'infC, iris3b, rpmI'–'lacZ*	this work
λMQ21-84ΔNB	*cI857, nin5; 'infC, iris84, rpmI'–'lacZ*	this work
λMQ21-12ΔNB	*cI857, nin5; 'infC, iris12, rpmI'–'lacZ*	this work
λMQ21-92ΔNB	*cI857, nin5; 'infC, iris92, rpmI'–'lacZ*	this work
λMQ21-162ΔNB	*cI857, nin5; 'infC, rpmI162a'-'lacZ*	this work
λMQ21-ΔApaLIΔNB	*cI857, nin5; 'infCΔApaLI, rpmI'–'lacZ*	this work
λMQ21-331infCΔNB	*cI857, nin5; 'infC331, rpmI'–'lacZ*	this work
λMQ21-332infCΔNB	*cI857, nin5; 'infC332, rpmI'–'lacZ*	this work
λMQ21-334'infCΔNB	*cI857, nin5; 'infC334', rpmI'–'lacZ*	this work
λMQ21-335infCΔNB	*cI857, nin5; 'infC335, rpmI'–'lacZ*	this work
λMQ21-336infCΔNB	*cI857, nin5; 'infC336, rpmI'–'lacZ*	this work
λMQ21-337infCΔNB	*cI857, nin5; 'infC337, rpmI'–'lacZ*	this work
λMQ21-76SIRΔNB	*cI857, nin5; 'infC, iris76, rpmI'–'lacZ*	this work
λMQ21-77SIRΔNB	*cI857, nin5; 'infC, iris77, rpmI'–'lacZ*	this work
λMQ21-77'SIRΔNB	*cI857, nin5; 'infC, iris77', rpmI'–'lacZ*	this work
λMQ21-78SIRΔNB	*cI857, nin5; 'infC, iris78, rpmI'–'lacZ*	this work
λMQ210-79SIRΔNB	*cI857, nin5; 'infC, iris79, rpmI'–'lacZ*	this work
λMQ21-80'SIRΔNB	*cI857, nin5; 'infC, iris80', rpmI'–'lacZ*	this work
λMQ21-81SIRΔNB	*cI857, nin5; 'infC, iris81, rpmI'–'lacZ*	this work
λMQ21-82SIRΔNB	*cI857, nin5; 'infC, iris82, rpmI'–'lacZ*	this work
λMQ21-(111/76SIR)ΔNB	*cI857, nin5; 'infC111, iris76, rpmI'–'lacZ*	this work
λMQ21-(111/77SIR)ΔNB	*cI857, nin5; 'infC111, iris77, rpmI'–'lacZ*	this work
λMQ21-(331infC/82SIR)ΔNB	*cI857, nin5; 'infC331, iris82, rpmI'–'lacZ*	this work
λMQ21-(332infC/81SIR)ΔNB	*cI857, nin5; 'infC332, iris81, rpmI'–'lacZ*	this work
λMQ21-(334'infC/79SIR)ΔNB	*cI857, nin5; 'infC334', iris79, rpmI'–'lacZ*	this work
λMQ21-(335infC/78SIR)ΔNB	*cI857, nin5; 'infC335, iris78, rpmI'–'lacZ*	this work
λMQ21-(336infC/77'SIR)ΔNB	*cI857, nin5; 'infC336, iris77', rpmI'–lacZ*	this work
λMQ21-(337infC/76SIR)ΔNB	*cI857, nin5; 'infC337, iris76, rpmI'–'lacZ*	this work
Plasmids		
pBR322	*amp; tet*	Bolivar *et al.* (1977)
pBL6	*tet; 'rpmI, rplT*	Lesage *et al.* (1992)
pTI3	*amp; 'thrS, infC, rpmI, rplT*	Lesage *et al.* (1992)

A prime on one side of a gene means that it is interrupted on that side.

to 1.24 pmol) was mixed with 16S rRNA from *Lactococcus lactis* (1 μg) as a carrier. Renaturation of RNA and chemical reaction conditions were essentially as described (Lesage *et al.*, 1992) with the following modifications. Reaction time with DMS was lowered to 3 min. CMCT was employed at a final concentration of 5 mg/ml and reaction time was 10 min in 'native buffer' (in the presence of magnesium) and 5 min in 'semi-native buffer' (in the absence of magnesium). Kethoxal was used for 5 min at a final concentration of 1 mg/ml. Detection of modified nucleotides and dideoxy sequencing of unmodified 'long transcript' were carried out as described (Lesage *et al.*, 1992). Sequences of the synthetic oligonucleotide primers are as follows: (DEL2) 5′ GTCTTTCACGCG-ATTAAGCACTTCC 3′ is complementary to positions 450–426 in *infC* except for a mismatch at position 438; (rpmI84) 5′ GGTCAGAATGTG-ACGCAG 3′ is complementary to positions 102 to 85 in *rpmI*.

Secondary structure modelling was done using Mulfold program (version 2.0 with suboptimal folding) which predicts RNA secondary structure by free energy minimization (Zuker, 1989). RNA secondary structures were drawn using the LoopViewer program (D.G.Gilbert, 1990. A Macintosh program published on the Internet, available via anonymous ftp to iubio.bio.indiana.edu.). Experimental data were used to alter the calculated secondary structure by prohibiting base pairing of nucleotides that are particularly reactive to chemicals. The first MulFold runs were made with a limited number of base pairing prohibitions (with the completely unreactive nucleotides) and the resulting folding compared with the experimental data. More constraints were introduced in the next runs until a better fit with the experimental data was obtained. Final changes in the secondary structure were done using only the experimental data.

Calculations of free energy for helix formation were made using thermodynamic parameters derived from the nearest-neighbour model (Freier *et al.*, 1986).

Acknowledgements

We thank Ciaran Condon, Jan van Duin and Marc Dreyfus for critical reading of the manuscript, and Maarten de Smit for helpful discussions. We thank Jacques Dondon and Monique Graffe for help with the initial characterization of some of the mutants. We are indebted to Christiane Branlant for communication of unpublished results. This work was supported by grants from the C.N.R.S. (UPR 9073), from the Ministère de l'Enseignement Supérieur et de la Recherche et de la Technologie and from EEC Human Capital and Mobility Network.

References

Brunel,C. *et al.* (1992) The domains of the *E.coli* threonyl-tRNA synthetase translational operator and their relation to threonine tRNA isoacceptors. *J. Mol. Biol.*, **227**, 621–634.

Burke,J.M. Belfort,M., Cech,T.R., Davies,R.W., Schweyen,R.J., Shub,D.A., Szostak,J.W. and Tabak,H.F. (1987) Structural conventions for group I introns. *Nucleic Acids Res.*, **15**, 7217–7221.

Butler,J.S., Springer,M., Dondon,J., Graffe,M. and Grunberg-Manago,M. (1986) *Escherichia coli* protein synthesis initiation factor IF3 controls its own gene expression at the translational level *in vivo*. *J. Mol. Biol.*, **192**, 767–780.

Butler,J.S., Springer,M. and Grunberg-Manago,M. (1987) AUU to AUG mutation in the initiator codon of the translation initiation factor IF3 abolishes translational autocontrol of its own gene (*infC*) *in vivo*. *Proc. Natl Acad. Sci. USA*, **84**, 4022–4025.

Christensen,T., Johnsen,M., Fiil,N.P. and Friesen,J.D. (1984) RNA secondary structure and translation inhibition: analysis of mutants in the *rplJ* leader. *EMBO J.*, **3**, 1609–1612.

Coetzee,T., Herschlag,D. and Belfort,M. (1994) *Escherichia coli* proteins, including ribosomal protein S12, facilitate *in vitro* splicing of phage T4 introns by acting as RNA chaperones. *Genes Dev.*, **8**, 1575–1588.

Das,A. (1993) Control of transcription termination by RNA-binding proteins. *Annu. Rev. Biochem.*, **62**, 893–930.

Davis,R.W., Botstein,D. and Roth,J.R. (1980) *Advanced Bacterial Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Freier,S.M., Kierzeck,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) Improved free-energy parameters for RNA duplex stability. *Proc. Natl Acad. Sci. USA*, **83**, 9373–9377.

Friesen,J.D., Tropak,M. and An,G. (1983) Mutations in the *rplJ* leader of *Escherichia coli* that abolish feedback regulation. *Cell*, **32**, 361–369.

Gallinaro,H., Domenjoud,L. and Jacob,M. (1994) Structural study of the 5′ end of a synthetic premessenger RNA from adenovirus. *J. Mol. Biol.*, **240**, 205–225.

Gutell,R.R. (1993) Comparative studies of RNA: inferring higher-order structure from patterns of sequence variation. *Curr. Opin. Struct. Biol.*, **3**, 313–322.

Haas,E., Morse,S.D.P., Brown,J.W., Schmidt,F.J. and Pace,N.R. (1991) Long-range structure in ribonuclease P RNA. *Science*, **254**, 853–856.

Hartz,D., McPheeters,D.S. and Gold,L. (1989) Selection of the initiator tRNA by *E.coli* initiation factors. *Genes Dev.*, **3**, 1899–1912.

Inoue,T. and Cech,T.R. (1985) Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc. Natl Acad. Sci. USA*, **82**, 648–652.

Jack,A., Ladner,J.E. and Klug,A. (1976) Crystallographic refinement of yeast phenylalanine transfer RNA at 2.5 Å resolution. *J. Mol. Biol.*, **108**, 619–649.

Jacquier,A. and Michel,F. (1987) Multiple exon-binding sites in class II self-splicing introns. *Cell*, **50**, 17–29.

Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.

Landick,R. and Yanofsky,C. (1987) Transcription attenuation. In Ingraham,J.L., Low,K.B., Magasanik,B., Schaechter,M., Umbarger, H.E. and Neidhardt,F.C. (eds), *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology.* American Society for Microbiology, Washington DC, Vol. 2, pp. 1276–1301.

Lesage,P. (1991) Regulation de l'Expression Génétique au Niveau de la Traduction dans l'Opéron infC-rpmI-rplT chez *E.coli*. Université de Paris XI.

Lesage,P., Truong,H.N., Graffe,M., Dondon,J. and Springer,M. (1990) Translated translational operator in *Escherichia coli*: autoregulation in the *infC-rpmI-rplT* operon. *J. Mol. Biol.*, **213**, 465–475.

Lesage,P., Chiaruttini,C., Dondon,J., Graffe,M., Milet,M. and Springer,M. (1992) Messenger RNA secondary structure and translational coupling in the *E.coli* operon encoding translation initiation factor IF3 and the ribosomal proteins, L35 and L20. *J. Mol. Biol.*, **228**, 366–386.

Michel,F. and Westhof,E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.*, **216**, 585–610.

Miller,E.S., Karam,J.D. and Spicer,E. (1994) Control of translation initiation: mRNA structure and protein repressors. In Karam,J.D. (ed.), *The Molecular Biology of Bacteriophage T4*. American Society for Microbiology, Washington DC, pp. 193–205.

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

Moazed,D., Stern,S. and Noller,H.F. (1986) Rapid chemical probing of conformation in 16S ribosomal RNA and 30S ribosomal subunits using primer extension. *J. Mol. Biol.*, **187**, 399–416.

Moine,H., Romby,P., Springer,M., Grunberg-Manago,M., Ebel,J.P., Ehresmann,B. and Ehresmann,C. (1990) *E.coli* threonyl-tRNA synthetase and $tRNA^{Thr}$ modulate the binding of the ribosome to the translation initiation site of the *thrS* mRNA. *J. Mol. Biol.*, **216**, 299–310.

Peattie,D.A. (1979) Direct chemical method for sequencing RNA. *Proc. Natl Acad. Sci. USA*, **76**, 1760–1764.

Philippe,C., Portier,C., Mougel,M., Grunberg-Manago,M., Ebel,J.P., Ehresmann,B. and Ehresmann,C. (1990) Target site of *E.coli* ribosomal protein S15 on its mRNA. Conformation and interaction with the protein. *J. Mol. Biol.*, **211**, 415–426.

Philippe,C., Eyermann,F., Bénard,L., Portier,C., Ehresmann,B. and Ehresmann,C. (1993) Ribosomal protein S15 from *E.coli* modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc. Natl Acad. Sci. USA*, **90**, 4394–4398.

Pleij,C.W.A. and Bosch,L. (1989) *Methods Enzymol.*, **180**, 289–303.

Pleij,C.W.A., Rietweld,K. and Bosch,L. (1985) A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.*, **13**, 1717–1731.

Quigley,G.J., Wang,A.H.J., Seeman,N.C., Suddath,F.L., Rich,A., Sussman,J.L. and Kim,S.H. (1975) Hydrogen bonding in yeast phenylalanine transfer RNA. *Proc. Natl Acad. Sci. USA*, **72**, 4866–4870.

Sacerdot,C., Chiaruttini,C., Engst,K., Graffe,M., Milet,M., Mathy,N., Dondon,J. and Springer,M. (1996) The role of the AUU initiation codon in the negative feedback regulation of the gene for translation initiation factor IF3 in *E.coli*. *Mol. Microbiol.*, in press.

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

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Spedding,G., Gluick,T.C. and Draper,D. (1993) Ribosome initiation complex formation with the pseudoknotted α operon mRNA. *J. Mol. Biol.*, **229**, 609–622.

Springer,M. (1996) Translational control of gene expression in *E.coli* and bacteriophage. In Lin,E.C.C. and Simon Lynch,A. (eds), *Regulation of Gene Expression in E.coli and Bacteriophage*. R.G.Landes Company, Austin, TX, pp. 85–126.

Springer,M. *et al.* (1985) Autogenous control of *Escherichia coli* threonyl-tRNA synthetase expression *in vivo*. *J. Mol. Biol.*, **185**, 93–104.

Springer,M., Graffe,M., Butler,J.S. and Grunberg-Manago,M. (1986) Genetic definition of the translational operator of the threonine tRNA ligase gene in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **83**, 4384–4388.

Tang,C.K. and Draper,D.E. (1989) Unusual mRNA pseudoknot structure is recognized by a translational repressor. *Cell*, **57**, 531–536.

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

Wertheimer,S.J., Klotsky,R.-A. and Schwartz,I. (1988) Transcriptional patterns for the *thrS-infC-rplT* operon of *Escherichia coli*. *Gene*, **63**, 309–320.

Westhof,E., Dumas,P. and Moras,D. (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.*, **184**, 119–145.

Winter,R.B., Morrissey,L., Gauss,P., Gold,L., Hsu,T. and Karam,J. (1987) Bacteriophage T4 regA proteins binds to mRNAs and prevents translation initiation. *Proc. Natl Acad. Sci. USA*, **84**, 7822–7826.

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

Zengel,J.M. and Lindahl,L. (1994) Diverse mechanisms for regulating ribosomal protein synthesis in *E.coli*. *Prog. Nucleic Acid Res. Mol. Biol.*, **47**, 331–369.

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

Received on February 19, 1996; revised on May 10, 1996