Volume 16 Number 18 1988 Nucleic Acids Research

Secondary structure of the leader transcript from the *Escherichia coli* S10 ribosomal protein operon

Ping Shen, Janice M.Zengel and Lasse Lindahl

Department of Biology, University of Rochester, Rochester, NY 14627, USA

Received May 20, 1988; Revised and Accepted 18 August, 1988

ABSTRACT

Genetic analysis of the autogenous control of the S10 ribosomal protein operon of Escherichia coli has suggested that the secondary or tertiary structure of the leader transcript is important for this regulation. We have therefore determined the secondary structure of the leader by enzyme digestion and chemical modification. Our results suggest that the 172 base leader exists in two forms, differing only immediately upstream of the Shine-Dalgarno sequence of the first gene. We discuss the possibility that the equilibrium between these alternate structures is important for the L4-mediated regulation of translation of the S10 operon. We have also determined the structure of several mutant transcripts. Correlation of these structures with the regulatory phenotypes suggest that a hairpin about 50 bases upstream of the first gene is essential for the control of translation of the operon. Finally, our results show that a two base substitution in an eight base loop destabilizes the attached stem.

INTRODUCTION

The synthesis of most ribosomal proteins (r-proteins) in Escherichia coli is regulated by autogenous control. That is, one of the proteins encoded by a given r-protein operon is not only a structural component of the ribosome, but also a regulatory protein. When this protein accumulates in excess of the specific binding sites available on nascent rRNA, it specifically represses its own operon. In the S10 r-protein operon, the autogenous control is mediated by the product of the third gene, L4, which is a repressor of both transcription and translation (1-5). The regulation of transcription is due to L4-stimulated transcription termination (attenuation) in the leader (4,6). Translation is probably regulated at the level of initiation at the first ribosome binding site in the operon.

Our laboratory has previously isolated mutations in the S10 leader which eliminate autogenous control at the transcription level, the translation level, or simultaneously at both levels (5,6). The analysis of these mutants demonstrated that extensive regions of the 172 base leader are required for the autogenous control. For example, base substitutions as far as 66 bases upstream of the initiation codon of the first gene eliminate both transcription and translation control. Also, attenuation control is abolished by a deletion 100 bases upstream of the site of transcription termination (5; L. Lindahl, R.

H. Archer and J. M. Zengel, unpublished results). That such extensive parts of the leader are required for the L4-mediated control suggests that specific secondary or tertiary structures may be essential for the regulation. In fact, the phenotypes of a particular set of mutations manipulating a potential hairpin structure within the leader indicate that secondary structure is indeed important, at least for the translation control (5).

To further our understanding of the regulation of the S10 operon, we have determined the secondary structure of the first 200 bases of the transcript from the wild-type operon and from several mutants. Based on in vitro structure mapping data and computer analysis, we have derived two secondary structures for the 172 base S10 leader, differing only in the region just upstream of the Shine-Dalgarno sequence. We propose that the S10 leader may exist as an equilibrium between these alternate forms and that L4 might modulate this equilibrium and thereby inhibit translation initiation.

Our structure analysis confirms that a hairpin structure previously proposed to be important for L4 control (4,5) does indeed exist in the S10 leader. Furthermore, analysis of transcripts containing base changes in this hairpin structure indicates that loss of translation control by L4 correlates with changes in the secondary structure of this region. This conclusion is especially interesting since the affected region is 40 bases upstream of the Shine-Dalgarno sequence. Surprisingly, these experiments revealed that base substitutions within the loop of this structure affect the stability of the stem.

MATERIALS AND METHODS

Plasmid Construction

To make pT724 (Fig. 1) we isolated from pLL36 (4) a FnuDII-SstI fragment carrying the beginning of the S10 operon, starting two bases upstream of the transcription initiation site and ending within the Shine-Dalgarno region of the S10 gene. This fragment was inserted between the HincII and SstI sites of M13mp18 (7) to create phage derivative M62. We then added the 3' adjacent region of the S10 operon by inserting the SstI-EcoRI fragment containing the intact S10 gene and 5' half of the L3 gene (Fig. 1). All the cloned S10 operon DNA was then transferred from the resulting phage, M65, to a T7 promoter vector, pT7-2 (US Biochemical Corporation), using the HindIII site in the M13 linker and the EcoRI site in the L3 gene.

pIM68 was constructed by first inserting the HindIII-EcoRI fragment from M65 (described above) into pIBI21 (International Biotechnologies, Inc.), a T7 promoter vector carrying the origin of replication from f1 phage. The vector DNA between the T7 promoter and the beginning of the S10 leader was then deleted by site-directed mutagenesis (8) using a 30-mer oligonucleotide consisting of the 15 base sequences on either side of the desired deletion.

Plasmid pIM62 was constructed by digesting pIM68 with SnaBI (which cuts in the middle of the S10 leader) and EcoRI. The created gap was then filled with the SnaBI-EcoRI fragment from phage M62 (described above). Mutant versions of pIM62 were constructed by replacing the SnaBI-SstI fragment with the corresponding fragment from M13 phages with the indicated base substitutions (5).

RNA labeling

RNA was synthesized in vitro using T7 RNA polymerase (US Biochemical Corporation) and a linearized template. Transcription reactions were incubated at 37°C for 1 hr and contained (in 50 μl): 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 10 mM dithiothreitol, 100 units RNasin ribonuclease inhibitor, 400 μM each of ATP, CTP, GTP, and UTP, and 120 units of T7 RNA polymerase. Transcripts were labeled at the 5' end by adding 200 μCi $\gamma[^{32}P]$GTP or labeled throughout by adding 10 μCi $[^{35}S]$UTP. The in vitro synthesized RNA was purified by extraction with phenol and chloroform-isoamylalcohol (24:1).

For in vivo labeling of RNA, a culture of E.coli strain LL308 (1) containing plasmid pLF1 harboring the 5' end of the S10 operon (9) was grown in MOPS medium (10) containing 0.2 mM phosphate. At about 10^8 cells per ml, 1.5 mCi $^{32}PO_4$ was administered to 3 ml of culture for 3 minutes. The cells were then lysed as described previously (3) and extracted with phenol and chloroform:isoamyl alcohol (24:1).

Digestions of RNA

Secondary structure analysis by partial enzyme digestion was done essentially as described (11,12). About $2x10^6$ cpm of 5' end-labeled RNA or 40 ng of unlabeled transcript was preincubated in 3 μl digestion buffer for 10 min at 37°C, after which 1 μl enzyme was added and incubation continued for 5 min at 37°C. The final reaction was 4 μl and contained 200 mM NaCl, 10 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5, 2.5 mg/ml yeast RNA, and an amount of RNase V1 or RNase T2 which had been determined experimentally to digest only a small fraction (estimated to be less than 25%) of the radioactive RNA. For RNase V1, that corresponded to 1 μl of a 1:100 or 1:200 dilution of the enzyme stock generously donated by Dr. J. Vournakis (units not determined). For RNase T2 (Sigma), 5-7.5 units were added.

Sequence markers were produced from 5' end-labeled RNA essentially as described previously (13). About $2x10^6$ cpm of RNA was partially digested at 60°C for 5 min in a 4 μl reaction mixture containing 20 mM Na-citrate pH 3.5, 10 mM EDTA, 7.5 M urea, 2.5 mg/ml yeast RNA, and 0.5-1 unit of RNase T1 or 0.5-1 unit of RNase U2. Similar aliquots of RNA were hydrolyzed partially in 4 μl 50 mM $Na_2CO_3/NaHCO_3$ (pH9.0) containing 5 mg/ml tRNA at 90°C for 1 min, 3 min, or 5 min.

Dimethyl sulfate modification

The procedure was adapted from published methods (14). About 40 ng RNA was incubated for 5 min at 37°C with 1 μl dimethyl sulfate or at 2°C for 90 min with 1 μl dimethyl sulfate diluted 1:6 in ethanol. Reactions (50 μl) contained 80 mM K-HEPES pH7.8, 300 mM KCl, and 20 mM $MgCl_2$, and were stopped by adding 20 μl of a solution containing 0.1 M Tris-acetate pH 7.0, 1.5 M Na-acetate and 0.1 mM EDTA.

Primer Extension

About 1 ng ($2x10^5$ cpm) 5' $[^{32}PO_4]$end-labeled primer (15) was mixed with approximately 40 ng RNA in a 7 μl reaction containing 50 mM KCl and 50 mM Tris-HCl pH 7.5. The mixture was incubated at 90°C for 1 min and then cooled slowly to room temperature. The primer was extended in a 5 μl reaction containing 1.5

μl annealing mixture, 0.5 μl 10x reaction buffer (250 mM Tris-HCl pH7.5, 250 mM KCl, 50 mM $MgCl_2$), 0.3 μl 0.1 M dithiothreitol, 0.5 μl deoxytriphosphate mixture (1 mM each of dATP, dCTP, dGTP, dTTP), and 0.5 μl (5 units) reverse transcriptase. Sequencing reactions also contained 15 μM of one of the four dideoxy ribonucleotides. After incubation at 45°C for 30 min, 2 μl of a solution containing 2 mM of each of the four dNTPs was added and the incubation was continued for another 15 min. The experiment in Fig. 3 was done with non-radioactive primer; the DNA was labeled by adding 10 μCi [^{35}S]dATP.

Gel electrophoresis

RNA fragments were electrophoresed at 40 V/cm through 0.3 mm thick 8% polyacrylamide gels containing 8 M urea, 89 mM Tris-borate (pH 8.3) and 2.5 mM EDTA.

Analysis of transcripts by hybridization and RNase protection

Radioactive transcripts were analyzed essentially as described (16,17). For analysis of in vivo labeled RNA we used an amount of RNA originating from about 0.5 ml cells. For analysis of in vitro synthesized RNA we used approximately 5 ng. In addition to the radioactive RNA, each hybridization reaction (20 μl) contained 25 μg yeast RNA, 1 μg of single-stranded DNA from phage M62 (see above) and 4 μl 5x hybridization buffer [3.75 M NaCl, 0.25 M HEPES-HCl (pH 7.0) and 5 mM EDTA]. The reaction was incubated at 68°C for 10 minutes, cooled to 50°C and incubated an additional 2 hrs. After chilling on ice, 160 μl T1 buffer [10 mM HEPES-HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA] containing 30 units RNase T1 was added and incubation continued for 30 minutes at 30°C. The RNAse was then inactivated with proteinase K (0.2 μg/ml) at 37°C for 30 min. The mixture was chilled and filtered slowly through a 25 mm S&S BA85 nitrocellulose filter which was then rinsed with 1 ml T1 buffer (without RNase T1). RNA was extracted from cut-up filters by boiling for 5 min in 0.5 ml 5 mM EDTA containing 10 μg yeast RNA.

RESULTS

Synthesis of S10 leader transcripts

Transcripts for secondary structure mapping of the S10 leader were synthesized in vitro with T7 RNA polymerase. To obtain templates for such transcription, we first constructed the plasmid pT724 (Fig. 1) containing a T7 promoter adjacent to a portion of the S10 operon beginning immediately upstream of the initiation site used by the E. coli RNA polymerase (see below) and ending in the second gene. T7 RNA polymerase transcripts from this plasmid contain 30 vector-encoded bases at the 5' end before the actual S10 leader transcript begins. Because we were concerned that these additional bases might affect the structure of the S10 leader transcript, we constructed a deletion derivative of pT724, called pIM68 (Fig. 1), lacking the unwanted vector DNA between the three G's of the natural T7 promoter and three G's at the 5' end of the S10 leader (Fig. 2). Since pIM68 does not contain a normal T7 promoter sequence, we were not certain of the start site used by T7 RNA polymerase for transcription of this template. We therefore determined the sequence of the 5' end of the transcript by primer extension in the presence of dideoxy ribonucle-

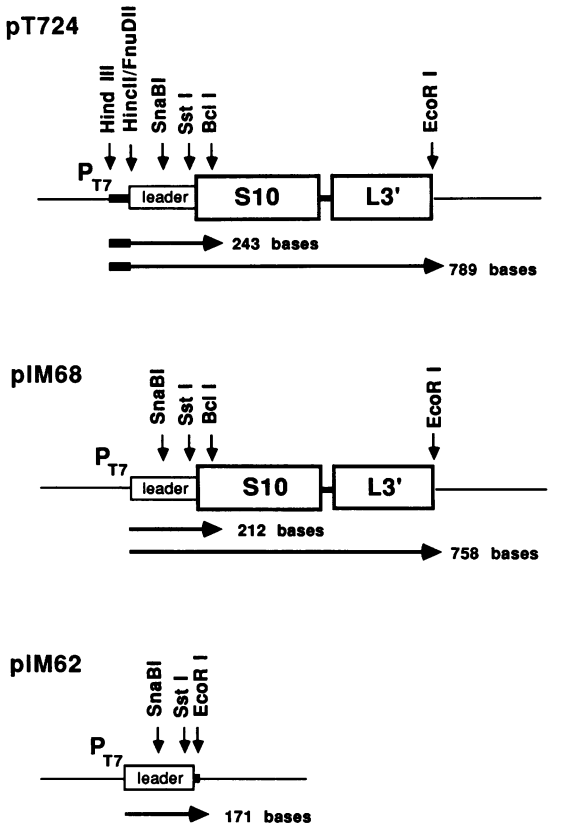


Figure 1. Maps of plasmids used as templates for in vitro transcription. Horizontal arrows indicate in vitro transcripts made by T7 RNA polymerase. The thickened regions of the horizontal arrows representing pT724 transcripts indicate the transcribed portion of the T7 promoter (P_{T7}) and linker region. Vertical arrows indicate positions of restriction sites used for construction of the plasmids (HindIII, HincII/FnuDII, SnaBI, SstI and EcoRI) or for truncating the templates (BclI and EcoRI sites). The SstI site overlaps with the Shine-Dalgarno region for the S10 gene.

otides. Gel electrophoresis of the products from these reactions showed bands consistent with the sequence of the 5' end of the S10 leader, as well as a strong band in all four tracks at the position corresponding to the most upstream of the three G's at the 5' end of the S10 leader (Fig. 3). These results demonstrate that the T7 RNA polymerase initiates transcription of pIM68 at the first of these three G's (Fig. 2).

We next wanted to compare the start site of the transcript from pIM68 with the start site used in vivo by E. coli RNA polymerase at the S10 promoter. The 5' end of the in vivo transcript was previously located to one of the three G's

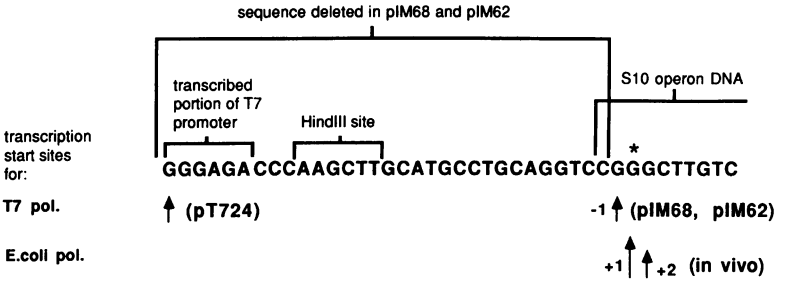


Figure 2. Summary of transcription initiation sites of the T7 and E. coli RNA polymerases. The sequence shown is that of pT724. Bases deleted to construct pIM68 and pIM62 are indicated. The HindIII site used to transfer the S10 leader sequence to the T7 promoter vector is also shown. The vertical arrows indicate the start sites determined for transcription by the T7 RNA polymerase in vitro (Fig. 3) or the E. coli RNA polymerase in vivo (Fig. 4). Position +1 (indicated by an asterisk) is the principal start site for in vivo transcription by the E. coli RNA polymerase (see text). The start site for T7 RNA polymerase on pT724 has not been determined experimentally, but is presumed correct because this plasmid carries the wild-type T7 promoter.

at the beginning of the leader (18). To define which of the three G's is the start site, we compared the 5' end of the in vivo transcript with the 5' end of the T7 RNA polymerase in vitro transcript by hybridization-RNase protection

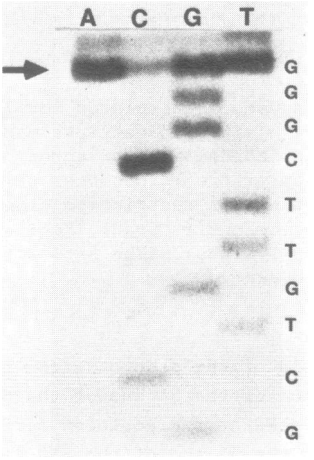


Figure 3. Sequencing of the 5' end of a transcript made in vitro from pIM68 by T7 RNA polymerase. An unlabeled primer complementary to the region between bases 13 and 25 was extended in the presence of [^{35}S] dATP and dideoxy nucleoside triphosphates as described in Materials and Methods. The lanes are identified by bases complementary to the incorporated dideoxy ribonucleotides to match the sequence in Figs. 2 and 7.

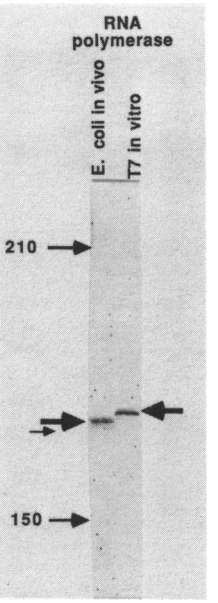


Figure 4. Mapping of the in vivo E.coli RNA polymerase start site in the S10 operon. Radioactive RNA was hybridized to single-stranded DNA from the M13 derivative M62 and then treated with RNase T1. The protected hybrids were collected on nitrocellulose filters, eluted, and electrophoresed through a denaturing 8% polyacrylamide gel. Left lane: $^{32}PO_4$ pulse-labeled RNA from E. coli strain LL308 carrying the plasmid pLF1. Right lane: ^{35}S-UTP labeled transcripts made in vitro from pIM68 with T7 RNA polymerase. Unmarked arrows indicate protected hybrids. Positions of RNA molecular weight markers are indicated with arrows labeled with the length (in bases) of each marker.

(16,17). Radioactive transcripts synthesized in vivo by the E.coli RNA polymerase or in vitro by the T7 RNA polymerase were hybridized to single-stranded DNA from phage M62 (see Materials and Methods) which contains almost the entire S10 leader, beginning one base upstream of the three G's (C -2) and ending at the Shine-Dalgarno region of the S10 gene. Non-hybridized parts of the transcripts were digested by T1 nuclease and the protected RNA fragments were analyzed on a denaturing polyacrylamide gel (Fig. 4). Since both the in vivo and in vitro transcripts extend beyond the downstream end of the insert in M62 and are therefore cleaved by T1 nuclease at precisely the same point at the 3' end, variations in the size of the protected fragments must be due to differences at the 5' ends. For the transcript made by the T7 RNA polymerase only one significant protected fragment was observed, corresponding to a transcript initiated at the first of the three G's at the beginning of the S10 leader. The E. coli in vivo transcript gave two protected fragments running one and two bases faster than the protected fragment from the T7 polymerase transcript. This

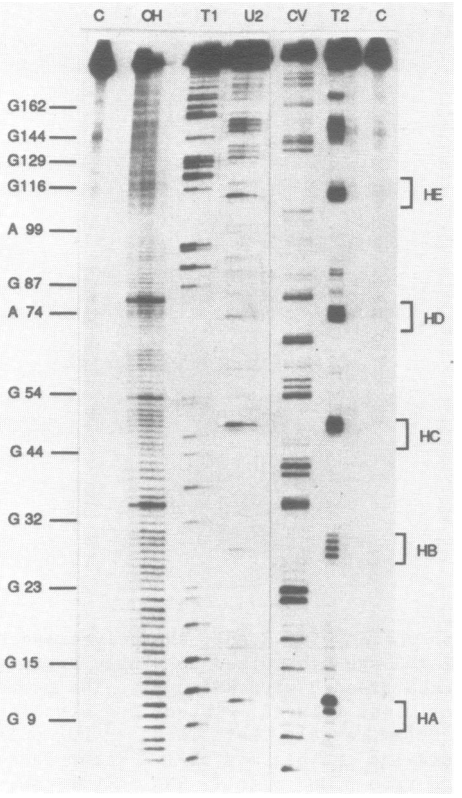


Figure 5. Enzymatic structure mapping of 5' end-labeled transcript from pT724. Aliquots of the transcript were partially degraded under various conditions as explained in Materials and Methods. C: RNA incubated in the buffer used for secondary structure analysis without enzyme. OH: Alkali hydrolysis products. T1 and U2: Products after partial digestion with RNase T1 and RNase U2, respectively, under denaturing conditions (sequence tracks). CV and T2: Products after partial digestion with RNase V1 and RNase T2, respectively, under non-denaturing conditions (tracks for structure mapping). G's and A's in the sequence as determined from the T1 and U2 tracks are indicated on the left. The brackets on the right indicate the loops of the hairpin structures discussed in the text.

experiment therefore suggests that the E. coli RNA polymerase has a major start at the second G and a minor start at the third G. We conclude that the transcript made by the T7 RNA polymerase from pIM68 begins only one base upstream of the principal start site used in vivo by the E. coli RNA polymerase (Fig. 2).

To obtain defined 3' ends on the in vitro transcripts, the plasmids pT724 and pIM68 were digested with restriction nucleases BclI or EcoRI (Fig. 1),

generating termination sites at the beginning of the S10 gene or within the L3 gene, respectively. A third plasmid, pIM62, containing only the S10 leader sequence and no structural genes of the S10 operon, was linearized with EcoRI (Fig. 1).

Structure mapping of the S10 leader

Our first approach to analyzing the secondary structure of the S10 leader transcript was to partially digest 5′ end-labeled transcripts with the single-strand specific T2 RNase (19) and the double-strand specific cobra venom RNase V1 (20). Fractionation of the products on denaturing polyacrylamide gels (Fig. 5) revealed a repeated motif of bands produced by RNase T2 cleavage at several consecutive bases bracketed by bands generated by RNase V1, suggesting the presence of stem-loop structures. We call these five hairpin structures HA through HE, named in the direction from the 5′ to the 3′ end (Fig. 5).

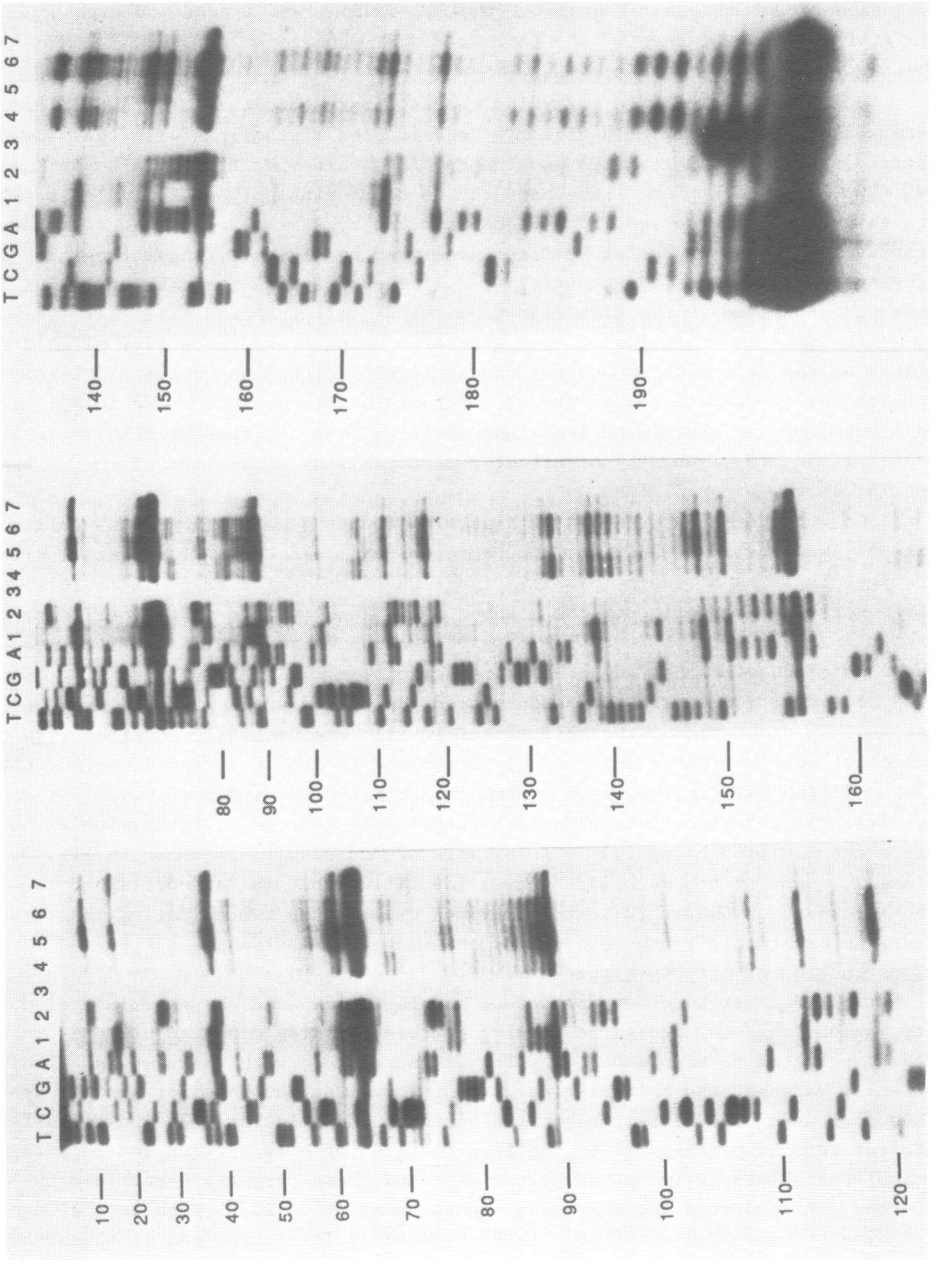
The data shown in Fig. 5 were generated using transcripts from pT724 truncated at the BclI site. However, the pattern of digestion in the S10 leader region was the same irrespective of which of the transcripts shown in Fig. 1 was used for the experiment (data not shown). Thus, it appears that the structure of the S10 leader is not affected detectably by extensions of the transcript at the 5′ or 3′ ends.

No cuts beyond approximately base +145 could be mapped accurately using the 5′ end-labeled transcript. We therefore examined RNase V1 sensitive sites further from the 5′ end by partial digestion of unlabeled transcripts followed by reverse transcription of the cleaved RNA. Results from this experiment, shown in Fig. 6, lanes 3 through 6, extend and confirm the RNase V1 cleavage data obtained with the end-labeled transcripts.

For additional structure information, unlabeled transcripts synthesized from pIM68 truncated at the BclI site (Fig. 1) were partially modified with dimethyl sulfate (DMS), which methylates bases in single-stranded regions (21). The modified RNA was then used as template for reverse transcription. Since reverse transcriptase is unable to transcribe Me-A or Me-C, it terminates before the modified bases (22). An example of the results is shown in Fig. 6, lane 2. Many of the methylated bases identified in this way correspond to positions cut by RNase T2 in the previous experiment, consolidating the structure information already obtained by the enzymatic analysis.

Construction of structure models

To construct secondary structure models we compared the experimental data to various computer-generated folding patterns, making minor adjustments to these patterns to accommodate results not compatible with the computer predictions. We generated a folding pattern of the first 200 bases of the S10 operon sequence using the classic algorithm of Zuker and Stiegler (23) and two different sets of parameters for the free energies of base pairing and loop formation: the values estimated by Salser (24) and those recently determined by Turner and coworkers for RNA helix formation at 37°C (25). Both sets of energy rules gave essentially the same folding pattern. This structure is illustrated in Fig. 7A. It has a calculated free energy of about -46 kcal/mol using the parameters of Freier et al. (25) disregarding contributions from the unpaired


T C G A 1 2 3 4 5 6 7
140
150
160
170
180
190
T C G A 1 2 3 4 5 6 7
80
90
100
110
120
130
140
150
160
T C G A 1 2 3 4 5 6 7
10
20
30
40
50
60
70
80
90
100
110
120

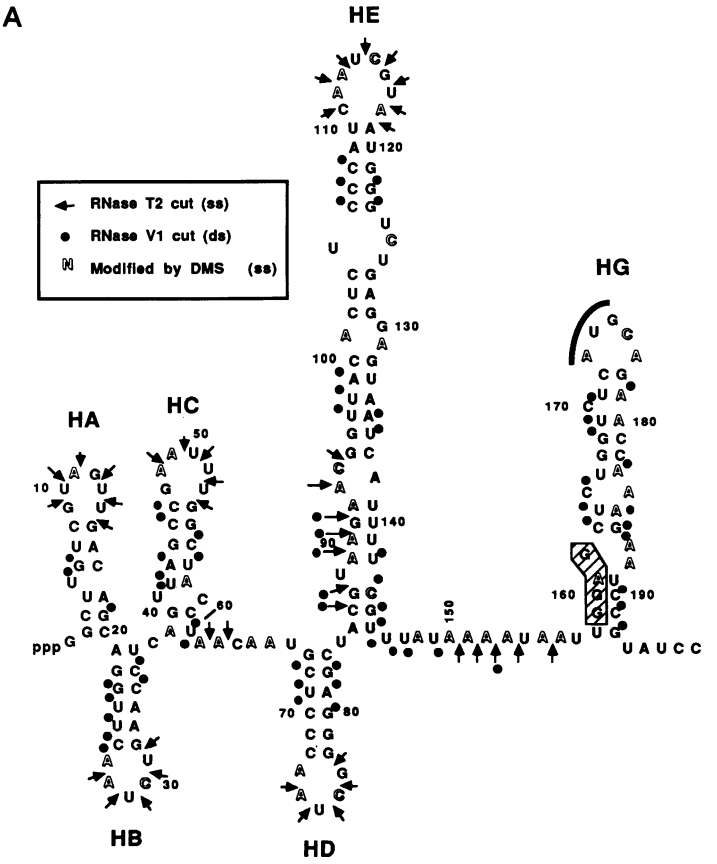
bases at the end of helices. The contributions of individual hairpins are given in the legend to Fig. 7. The model contains all five of the hairpin structures identified by the enzymatic structure mapping in Fig. 5 and has one additional hairpin (called HG) containing the Shine-Dalgarno region and the initiation codon of the S10 gene. All six hairpins including their bulges are supported by the experimental results (Fig. 7A). Likewise, most of the single-stranded regions connecting the stem-loops are compatible with the data. However, two regions are ambiguous. One of these is between positions 87 and 93, for which the structure mapping data generated evidence for both single- and double-strandedness. This region is part of an imperfect stem at the bottom of HE, suggesting that the base pairing in this region could be unstable. Alternatively, the region could be involved in two different structures which are in equilibrium with each other. The other region of ambiguity is around base 153, which also generated unclear structure data and will be discussed further below.

The computer program used above only displays the folding pattern which has the minimal calculated free energy. Given the uncertainties which still exist in the free energy values for base pairing and loop formation, it would be desirable to inspect alternative folding patterns with calculated ΔG values less favorable than the value of the optimal solution. A program which allows the user to view any possible folding pattern, irrespective of ΔG value, has recently been designed by Zuker (26) and tested by J. Jaeger and D. Turner (personal communication). We used this program to view all the possible folding patterns of the S10 leader with ΔG values within 10% of the optimal solution. The energy parameters used were the 37°C values determined by Freier et al. (25), with additional recent refinements (D. Turner, personal communication). This procedure generated 21 different folding patterns, most of which were essentially identical to the model shown in Fig. 7A, differing only with respect to minor modifications of base pairing around some bulges. A few models differed radically in the regions of hairpins HA, HB and HC, but could be eliminated by comparison with our structure probing data.

A very interesting variation in the folding pattern was seen between bases 135 and 155, the region generating the ambiguous structure mapping information

Figure 6. Chemical and enzymatic structure mapping of unlabeled transcript from pIM68. Aliquots were incubated with DMS or RNase V1 as described in Materials and Methods. The modified or partially cleaved transcripts were then reverse transcribed using a primer complementary to bases 24-39 of the S10 gene. T, C, G, A: Reverse transcription of unmodified transcript in the presence of dideoxy ribonucleotides (sequence markers). The lanes are identified by bases complementary to the incorporated dideoxy ribonucleotides to match the sequence in Fig. 7. Bases are numbered according to the system in Fig. 7. 1: Mock sample for DMS modification. 2: DMS modification at 37°C. 3 and 4: RNase V1 (1:100 dilution) digestion for 5 and 1 min, respectively. 5 and 6: RNase V1 (1:200 dilution) digestion for 1 and 5 min, respectively. 7: Mock sample for RNase V1 digestions. The bands in the control lanes (1 and 7) are due to spontaneous stops and pauses in the reverse transcriptase reaction. The few differences between the patterns in the two lanes are apparently due to the different buffer conditions in the RNase V1 and DMS reactions.

previously pointed out. Varying numbers of the U's in the distal end of HE were eliminated from HE and base paired with A's in the region between bases 150 and 160. The most extreme of these models, shown in Fig. 7B, has a calculated ΔG value of -42 kcal/mol, or approximately 10% higher than the model in Fig. 7A. Given that a number of the parameters for calculation of free energy have not been tested experimentally (25,27), the difference between the calculated free energies of the two models is probably not significant. Virtually all the structure mapping information which does not fit the model in Fig. 7A is in agreement with the structure in Fig. 7B. However, some structure probing data are satisfied by Fig. 7A, but not by Fig. 7B. Thus, even though the model in Fig. 7A is compatible with the bulk of the structure mapping data, no single model is fully compatible with all structure mapping data. Our interpretation of this finding is that the S10 leader exists in an equilibrium between two or more structures. Because of the relative strength of the cutting with RNases T2 and V1 in the ambiguous regions, we suggest that the S10 leader exists mainly in the form shown in Fig. 7A under the conditions used in our experiments.



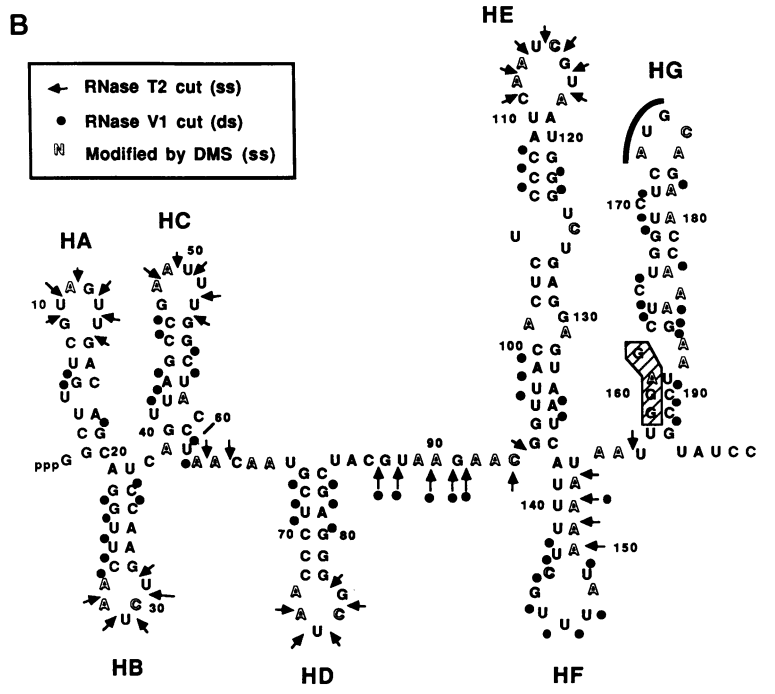


Figure 7. Models of the secondary structure of the leader of the S10 operon. The sites of RNase cleavages and DMS modifications observed in the experiments illustrated in Figs. 5 and 7, as well as several other similar experiments not shown, are summarized by the symbols explained in the boxes. HA through HG indicate hairpin structures discussed in the text. The calculated free energies (in kcal/mol) of individual hairpins in model A are: -3.7 (HA), -5.5 (HB), -5.7 (HC), -8.8 (HD), -15.2 (HE), and -7.1 (HG). The total ΔG of model A is -46 kcal/mol. In model B, the calculated free energy of the shortened hairpin HE is -11.5 kcal/mol and of HF +0.5 kcal/mol. The overall calculated ΔG of model B is -42 kcal/mol.

Structure mapping of mutant transcripts

We have previously analyzed the regulatory response to L4 oversynthesis in a series of mutants carrying two-base substitutions in the upper stem-loop of hairpin HE (see Discussion and ref.5). This part of the leader contains a 9 base sequence (bases 117 through 125) also found in 23S rRNA at the binding site for r-protein L4 (18). Furthermore, this region is immediately upstream of the site of L4-mediated transcription termination, which occurs around base 140 (4; J. M. Zengel and L. Lindahl, unpublished observations). The mutations had been designed to test the role of transcript structure in the L4-mediated regulation of the operon. One mutant RNA (the "stem mutant") contained a two base substitution in the ascending side of the upper stem (Fig. 8) and was expected to cause disruption of the hairpin structure. A second mutant RNA

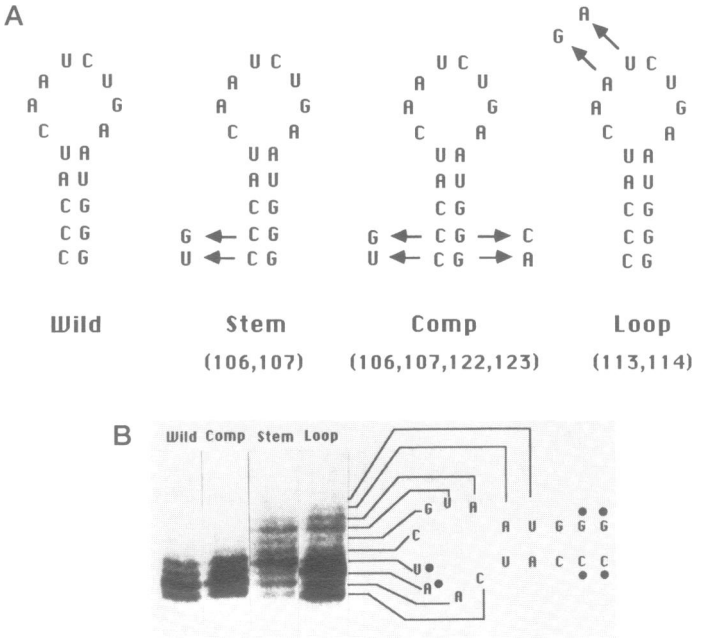


Figure 8. Enzymatic structure determination of transcripts with mutations in hairpin HE. A. The sequence of the loop and upper stem of the hairpin is shown for the wild-type sequence ("Wild") with the indicated base changes in the "Loop", "Stem" and Compensatory ("Comp") mutants. The positions of the base substitutions relative to the start of transcription are indicated in parentheses. B. End-labeled transcripts from derivatives of pIM62 containing the various base substitutions were partially digested with RNase T2. Only bands corresponding to the bases in the loop and upper stem of HE are shown, since no differences were seen in other parts of the molecule. The closed circles indicate the positions of relevant base substitutions.

(the "compensatory mutant") contained the same two base substitutions as well as two other base changes on the descending side of the upper stem which we expected would restore the basic stem-loop structure but with a different primary sequence. The last of the mutant RNAs (the "loop mutant") contained a two base substitution in the loop of the HE structure which, according to computer predictions, would not affect the secondary structure.

To determine if our predictions for the various mutant structures were correct, we used partial digestion with RNases T2 and V1 to map the structures of 5' end-labeled transcripts from derivatives of pIM62 carrying the various mutations. As expected, the stem mutation made the bases in the upper stem of HE accessible to cleavage with RNase T2 (Fig. 8). That is, the base substitutions disrupted the stem. The digestion patterns of the remaining parts of the mutant molecule were indistinguishable from the results of the wild-type tran-

script (data not shown), indicating that the disruption of the upper stem of HE is local and does not lead to major rearrangement of the RNA structure. The RNA molecule from the compensatory mutant also yielded the predicted result: changing the sequence on the other side of the stem to allow base pairing of the two sets of two-base substitutions restored the digestion pattern to that of the wild-type hairpin structure (Fig. 8).

A surprising result was obtained with a mutant changing bases 113 and 114 in the loop of HE. Since these two bases are not immediately adjacent to the HE stem, and thus probably do not co-stack with the bases in the stem, we had expected that the base changes would have no effect on the stability of the stem. However, we found that the loop mutation made several bases at the top of the stem accessible to digestion with RNase T2 (Fig. 8), indicating that the base changes in the loop in fact destabilize the stem.

As part of our mutant analysis, we had also constructed a mutant, called Δ72, removing bases 15 through 86 of the S10 transcript, i.e. most of the 5′ half of the leader. This mutation eliminated the L4-mediated attenuation, even though it left intact all but two of the bases constituting hairpin HE, including the sequence around the attenuation site (5). To see if the deletion might have indirectly altered the structure of hairpin HE, we analyzed the structure of the Δ72 mutant transcript made from a deletion derivative of pIM62. The RNase T2 and V1 digestion patterns of the RNA downstream of the Δ72 deletion were the same as we had previously observed for the wild-type transcript (Fig. 9), indicating that HE has virtually the same structure in the Δ72 transcript as in the wild-type transcript.

DISCUSSION

Secondary structure models

Extensive portions of the leader region of the S10 r-protein operon of *E. coli* are required for the autogenous control of both transcription attenuation and translation (5), suggesting that the secondary or tertiary structure of the leader plays a role in mediating the regulatory responses. In the present investigation, we have examined the secondary structure of the leader. By comparing structure mapping data with the 21 computer-generated folding patterns with the lowest calculated free energies, we conclude that there is one model, shown in Fig. 7A, which accounts for most of the experimental data. For about 90% of the leader sequence this model is unambiguous in the sense that all computer models which can satisfy the experimental observations include hairpins HA through HE and HG. However, it is possible that minor details such as the exact length of stems or the exact position of bulges could deviate slightly from the model shown in Fig. 7A.

There are two regions where the model in Fig. 7A does not fit the structure mapping data or where the diester bonds were cleaved by both RNase T2 and RNase V1. These ambiguous regions include bases 86 through 92 and 145 through 154. In fact, none of the 21 models we examined were totally compatible with all the experimental data from these regions. Hence, we propose that the leader structure is in equilibrium between two or more secondary structure forms

which are very similar but differ with respect to the ambiguous regions defined above. One plausible alternate structure is shown in Fig. 7B. Collectively the two structures in Fig. 7 can account for all data obtained with chemical modification and RNase T2 digestion and for all but two of the cleavages (after bases 146 and 147) observed with RNase V1. The proposed alternate structure (Fig. 7B) contains one hairpin (HF) not present in the model in Fig. 7A. Being composed of only AT base pairs, the HF hairpin is intrinsically very weak, but it might be stabilized by forming a coaxial helix with the lower stem of HE.

Structure of mutant transcripts

For most of the mutant RNA molecules we analyzed, the predicted secondary structure agreed with the actual experimental data. However, we were surprised to find that base changes in the loop of the HE hairpin structure disrupt base-pairing of the upper stem. The destabilizing effect of the HE loop mutation suggests that the 8 base loop is not entirely flexible. It appears to have a specific structure which can be perturbed by base substitutions, such that the altered shape of the mutated loop necessitates the breaking of base pairs at the top of the stem in order to maintain the continuity of the RNA backbone. The change of the loop shape could have two reasons. Perhaps the stacking interactions between bases in the loop are sufficiently strong to bestow a preferred conformation on the loop. For example, it is possible that the mutant loop has a different shape because it contains three consecutive purines, where the wild-type loop only has two. Alternatively, bases in the loop could base pair with another single-stranded region elsewhere in the molecule to generate a specific structure. However, the latter possibility seems less likely, since all diester bonds in the HE loop were accessible to the single-strand specific probes, but none were cleaved with double-strand specific V1 nuclease. Recently, it was proposed that loops containing CUUCGG are associated with exceptionally stable hairpins (28). None of the loops in the proposed structures of the S10 leader contain this motif, but it is possible that the sequence AAUPy found in four of the S10 loops also could be a stabilizing element in hairpin structures.

Transcript structure and the regulation of the S10 operon

The models in Fig. 7 are consistent with our previous genetic analysis of the autogenous control of the S10 operon, which suggested that a specific secondary structure in the upper portion of hairpin HE is essential for the L4-mediated control of translation (5). In these genetic studies, we observed that base changes on either side of the putative stem eliminated L4-mediated translation control. Furthermore, "compensatory" base changes which restored the stem-loop structure, but with a different primary sequence, also restored translation regulation. The mutant analysis also showed that the translation control could be disrupted by base substitutions in the HE loop, a result which we originally interpreted as an indication that a specific base sequence in the loop is critical to the regulatory mechanism. However, since the secondary structure analysis indicates that the upper stem of HE is disrupted at the top by the base substitutions in the loop, we cannot be sure if the phenotypes of the loop mutants are due to a direct role of the primary base sequence of the

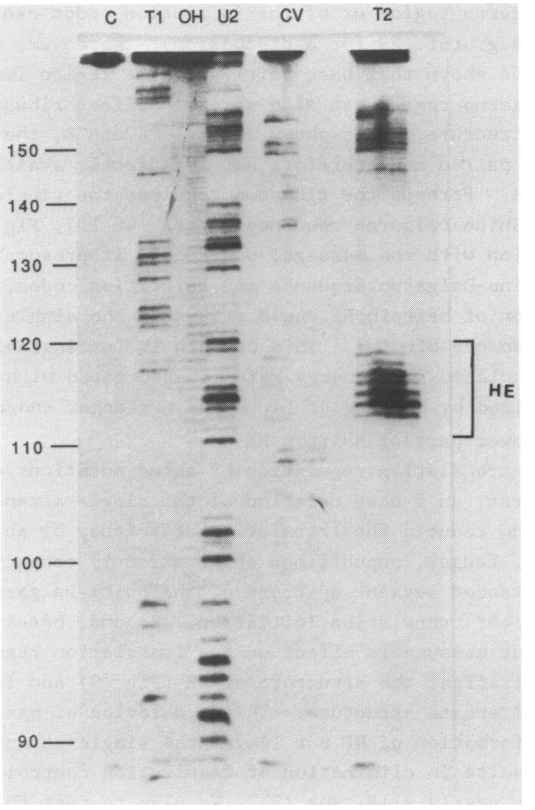


Figure 9. Enzymatic structure determination of the transcript from a mutant with a 72 base deletion in the 5' end of the leader. End-labeled transcript from a pIM62 derivative lacking bases 15 through 86 was partially digested with different RNases. C: RNA incubated in the buffer used for secondary structure analysis without enzyme. OH: Alkali hydrolysis products. T1 and U2: Products after partial digestion with RNase T1 and RNase U2, respectively, under denaturing conditions (sequence tracks). CV and T2: Products after partial digestion with RNase V1 and RNase T2, respectively, under non-denaturing conditions (tracks for structure mapping). The numbers on the left correspond to base numbers in the wild-type sequence (Fig. 7).

loop in translation control, or to a requirement for stable base-pairing within the adjacent stem structure. In any event, further experiments are required to clarify how base substitutions 60-70 bases upstream of the first structural gene of the operon can influence the L4-mediated regulation of translation initiation. The secondary structure models presented here suggest a working model for the translation control. Namely, translation control could be accomplished by an L4-mediated shifting of the proposed equilibrium between the structures shown in Figs. 7A and 7B. It has been established that base pairing

of the Shine-Dalgarno region or of the initiation codon can inhibit translation initiation (see e.g. ref. 29 for a discussion). Moreover, experiments with fd phage gene IX have shown that base pairing in the region immediately upstream of the Shine-Dalgarno region can also strongly affect ribosome binding to mRNA (30). In both structure models shown in Fig. 7A and B, the Shine-Dalgarno sequence is base paired and therefore not immediately available for interaction with the 16S rRNA. Perhaps the ribosome requires the single-stranded region upstream of the Shine-Dalgarno sequence (bases 146-157, Fig. 7A) to form its initial association with the message; once bound it presumably could gain access to the Shine-Dalgarno sequence and initiation codon, and begin translation. Formation of hairpin HF would sequester the single-stranded region and thereby block ribosome binding. This hairpin is intrinsically very weak (essentially no calculated free energy gain is associated with its formation), but it may be stabilized by binding of L4 to the messenger and/or by making a joint helix with the lower part of hairpin HE.

Analysis of translation regulation of three mutations agree with the proposed model. First, an 8-base deletion of the single-stranded region between hairpins HE and HG reduces the translation efficiency by about ten-fold (L. Lindahl and J. M. Zengel, unpublished observations), compatible with the notion that a single-stranded segment upstream of the Shine-Dalgarno sequence is required for efficient translation initiation. Second, bases 15 through 86 can be deleted without measurable effect on the translation regulation (5). This deletion does not affect the structure of HE (Fig. 9) and leaves possible the switch into an alternate structure. Third, deletion of bases 94 through 142, which precludes formation of HF but leaves the single-stranded region upstream of HG intact, results in elimination of translation control, but leaves translation efficiency nearly wild-type (5). We plan to test this hypothesis further by creating mutations in the S10 leader sequence which manipulate the stability of hairpin HF.

The secondary structure models offer less insight into the mechanism of L4-mediated attenuation. We know that L4 causes termination of transcription within the string of U's around base 140 of the leader (4; J. M. Zengel and L. Lindahl, unpublished experiments). Thus, hairpin HE appears to function as a rho-independent terminator whose strength is in some way modulated by L4. The secondary structure data support the existence of this terminator-like structure but cannot explain the role of L4. Moreover, even though leader RNA from the compensatory mutant assumes the same secondary structure as wild-type RNA, the compensatory base changes do not restore transcription control by L4 (5). On the other hand, the 113,114 loop mutation, which disrupts base pairing adjacent to the loop, leaves L4-mediated attenuation control intact. And finally, a deletion removing 72 bases from the 5' half of the S10 leader abolishes transcription control yet leaves essentially all the HE hairpin structure intact (Fig. 9), indicating that the sequences upstream of HE are also required for transcription control. At this point we have no simple model to account for all these observations. However, since we now have defined the structural

domains within the S10 leader, we can design mutations which delete or alter specific domains to test their contribution to this control mechanism.

Comparison of structures of S10 leader and 23S rRNA

The central feature in current models for autogenous control of r-protein synthesis is that the regulatory r-protein binds directly to its own messenger RNA, presumably via structural similarities between the mRNA leader and the appropriate binding site on rRNA. Indeed, specific binding of the regulatory r-protein has been demonstrated for the L10 (31) and alpha operons (32). Although we have not yet demonstrated a specific interaction between L4 and the S10 leader, we assume that L4 asserts its regulatory role by binding to a specific target in the S10 leader. However, we cannot see any homology between the secondary structures of the S10 leader and the regions around base 300 and base 600 of the 23S rRNA to which protein L4 has been cross-linked (33,34) and which therefore presumably constitute at least part of the L4 binding domain on 23S rRNA. There is an interrupted primary sequence homology consisting of several segments of 3 to 9 bases each of which are found both in the S10 leader and around base 600 of 23S rRNA (18). However, genetic experiments in our laboratory have failed to demonstrate any special regulatory role for these homology regions (4-6; J. M. Zengel and L. Lindahl, unpublished experiments). At this time it is therefore not clear whether the primary sequence homology is relevant for the interaction of L4 with its messenger. Furthermore, the cross-linking of L4 to two different regions of the 23S rRNA suggests that the homology, if it exists, may be found at the three dimensional level. Further characterization of the L4 interaction with rRNA and mRNA should be very interesting.

ACKNOWLEDGMENTS

We thank J. Vournakis and coworkers for introducing us to the methodology of RNA structure analysis. We are especially grateful to D. Turner and J. Jaeger for help with the computer analysis and many stimulating discussions. This work was supported by a research grant and a Research Career Development Award to LL from the National Institute for Allergy and Infectious Diseases.

REFERENCES

1. Lindahl, L., and Zengel, J. M. (1979) Proc. Nat. Acad. Sci. USA 76: 6542-6546.
2. Yates, J. L. and Nomura, M. (1980) Cell 21: 517-522.
3. Zengel, J. M., Mueckl, D. and Lindahl, L. (1980) Cell 21: 523-535.
4. Lindahl, L., Archer, R. H., and Zengel, J. M. (1983) Cell 33: 241-248.
5. Freedman, L. P., Zengel, J. M., Archer, R. H. and Lindahl, L. (1987) Proc. Nat. Acad. Sci. USA 84: 6516-6520.
6. Lindahl, L. and Zengel, J. M. (1988) in Genetics of Translation. Tuite, M., Picard, M. and Bolotin-Fukuhara, M., Eds., pp. 105-115, Springer-Verlag, New York.
7. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33: 103-119.
8. Kunkel, T. A. (1985) Proc. Nat. Acad. Sci. USA 82: 488-492.
9. Freedman, L. P., Zengel, J. M. and Lindahl, L. (1985) J. Mol. Biol. 185: 701-712.
10. Neidhardt, F. C., Bloch, P. L. and Smith, D. F. (1974) J. Bacteriol. 119: 736-747.

11. Vournakis, J. N., Celanto, J., Finn, M., Locard, R. E., Mitra, T., Pavlakis, G., Troutt, A., van den Berg, M. and Wurst, R. (1981) in Gene Amplification and Analysis, vol. 2. Chirikjian, J. G. and Papas, T. S., Eds., pp. 267-298, Elsevier North Holland, New York.
12. Wurst, R. M., Vournakis, J. N. and Maxam, A. M. (1978) Biochemistry 17: 4493-4499.
13. Donis-Keller, H., Maxam, A. M. and Gilbert, W. (1977) Nucleic Acids Res. 4: 2527-2538.
14. Stern, S., Wilson, R. C. and Noller H. F. (1986) J. Mol. Biol. 192: 101-110.
15. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, New York
16. Hansen, U. and Sharp, P. A. (1983) EMBO J. 2: 2293-2303.
17. Lamond, A. I. and Travers, A. A. (1985) Cell 40: 319-326.
18. Olins, P. and Nomura, M. (1981) Cell 26: 205-211.
19. Uchida, T. and Egami, F. (1967) in Methods of Enzymology XII. Grossman, L. and Moldave, K., Eds., pp. 228-247, Academic Press, New York.
20. Vassilenko, S. K. and Rythe, V. C. (1975) Biokhimiya 40: 578-582
21. Lawley, P. D. and Brooks, P. (1963) Biochem. J. 89: 117-138
22. Inoue, T. and Cech, T. R. (1985) Proc. Nat. Acad. Sci. USA 82: 648-652
23. Zuker, M. and Stiegler, P. (1981) Nucleic Acids Res. 9: 133-148
24. Salser, W. (1977) Cold Spring Harbor Symp. Quant. Biol. 42: 985-1002.
25. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. and Turner, D. (1986) Proc. Nat. Acad. Sci. USA 83: 9373-9377.
26. Zuker, M. (1987) in Mathematical Methods for DNA Sequences. Waterman, M. S., Ed., in press, CRC Press, Florida.
27. Turner, D. H. and Sugimoto, N. (1988) Ann. Rev. Biophys. and Biophys. Chem. in press.
28. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O. E., Tinoco, I., Brody, E. N. and Gold, L. (1988) Proc. Nat. Acad. Sci. USA 85: 1364-1368.
29. Berkhout, B., Schmidt, B. F., van Strien, A., van Boom, J., van Westrenen, J. and van Duin, J. (1987) J. Mol. Biol. 195: 517-524.
30. Blumer, K. J., Ivey, M. R. and Steege, D. A. (1987) J. Mol. Biol. 197: 439-451.
31. Johnsen, M., Christensen, T., Dennis, P. P. and Fiil, N. P. (1982) EMBO J. 1:999-1004.
32. Deckman, I. C. and Draper, D. E. (1985) Biochemistry 24: 7860-7865.
33. Maly, P., Rinke, J., Ulmer, E., Sweib, C. and Brimacombe, R. (1980) Biochemistry 19: 4179-4188.
34. Gulle, H., Hoppe, E., Osswald, M., Greuer, B., Brimacombe, R. and Stoeffler, G. (1988) Nucleic Acids Res. 16: 815-832