RNA-structural Mimicry in Escherichia coli Ribosomal Protein L4-dependent Regulation of the S10 Operon*

Ulrich Stelzl‡§, Janice M. Zengel, Marina Tovbina‡, Marquis Walker¶, Knud H. Nierhaus§, Lasse Lindahl¶, and Dinshaw J. Patel‡

From the ‡Memorial Sloan Kettering Cancer Center, Cellular Biochemistry and Biophysics Program, New York, New York 10021, the §Max-Planck-Institut für Molekulare Genetik, AG Ribosomen, Ihnestrasse 73, D-14195 Berlin, Germany, and the ¶Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, Maryland 21250

Ribosomal protein L4 regulates the 11-gene S10 operon in Escherichia coli by acting, in concert with transcription factor NusA, to cause premature transcription termination at a Rho-independent termination site in the leader sequence. This process presumably involves L4 interaction with the leader mRNA. Here, we report direct, specific, and independent binding of ribosomal protein L4 to the S10 mRNA leader in vitro. Most of the binding energy is contributed by a small hairpin structure within the leader region, but a 64-nucleotide sequence is required for the bona fide interaction. Binding to the S10 leader mRNA is competed by the 23 S rRNA L4 binding site. Although the secondary structure of the mRNA and rRNA binding sites appear different, phosphorothioate footprinting of the L4-RNA complexes reveals close structural similarity in three dimensions. Mutational analysis of the mRNA binding site is compatible with the structural model. In vitro binding of L4 induces structural changes of the S10 leader RNA, providing a first clue for how protein L4 may provoke transcription termination.

Ribosomal proteins are encoded on the bacterial chromosomes in multigene operons that facilitate stoichiometric production of the more than 50 ribosomal proteins (r-proteins).1 Expression of most of these operons is translationally regulated by a single regulatory r-protein encoded by a given polycistronic mRNA. This operon-specific autogenous control mechanism is elicited when repressor r-proteins are not consumed during the assembly of ribosomal subunits. In the absence of sufficient ribosomal RNA targets, the resulting excess of “free” repressor proteins bind to their own mRNAs and block translation of the polycistronic mRNA (reviewed in Refs. 1 and 2). R-protein L4 from Escherichia coli specifically regulates the S10 operon, which codes for 11 r-proteins including L4 itself (3, 4). L4 is unique among the regulatory ribosomal proteins because it regulates not only translation but also transcription of the S10 operon mRNA. The latter form of regulation is accomplished by L4 stimulation of transcription termination at a terminator (attenuator) structure in the mRNA upstream of the initiation codon of the first gene of the operon (5). Determinants necessary for L4-mediated autogenous control are contained within a 172-nucleotide 5′-untranslated region of the S10 operon mRNA. This region folds into six stem-loop structures termed helices HA, HB, HC, HD, HE, and HG (Fig. 1A) (6, 7).

The mRNA sequences necessary for L4-mediated transcription and translation control overlap, but are not identical (8, 9). Helix HE and the unstructured sequence immediately downstream of this hairpin are necessary for translational control, whereas helices HD and HE are required for transcriptional control (8, 10, 11). L4-stimulated transcription termination occurs within the U cluster (nt 139–149) (9) in helix HE, which resembles a Rho-independent terminator. In vitro transcription experiments (10, 12) suggest that the transcription attenuation process involves three ordered steps: (i) spontaneous pausing of the RNA polymerase at the attenuation site, (ii) NusA-mediated stabilization of the pause, and (iii) additional stabilization of the pause involving protein L4 that leads to transcription termination.

Binding of a regulatory r-protein to specific targets in both rRNA and mRNA is fundamental to the autogenous control model (13). The minimal rRNA target of L4, identified by serial selection of random E. coli rRNA fragments binding to L4 (14), is a four-helix junction composed of nucleotides 295–343 in Domain I of the 23 S rRNA. Binding of protein L4 to this RNA sequence generates a characteristic iodine cleavage pattern in the U321-loop of 23 S rRNA (15). Interestingly, the same region of 23 S rRNA competes with the paused transcription complex for L4, eliminating the ability of the protein to stimulate transcription termination (16). These results show that binding of 23 S rRNA to L4 inactivates the regulatory capacity of the protein and suggest that the 23 S rRNA and mRNA binding sites for L4 might share certain critical features. However, mutations in protein L4 can distinguish its ability to regulate the expression of the S10 operon from its ability to be incorporated into the ribosome (17).

Here we demonstrate direct binding of r-protein L4 to the S10 mRNA leader. Furthermore, we analyze the requirements in the RNA binding site for L4 binding. Footprinting results indicate structural changes of the leader RNA upon binding, and reveal structural similarities between the rRNA and mRNA L4 binding sites. A structure model for the L4-mRNA interaction is proposed.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—Protein L4 from E. coli (L4EC) was purified from 50 S subunits as described (18). The clone of the
full-length *Thermus aquaticus* L4 protein was a kind gift from Dr. Alexander Mankin (University of Illinois, Chicago, IL). Protein L4THD7 is a variant of *T. aquaticus* L4 with a deletion in which 46 amino acids (Thr<sup>88</sup> to Lys<sup>93</sup>, corresponding to Thr<sup>86</sup> to Arg<sup>91</sup> in *E. coli*, Fig. 1B) were replaced with a single glycine. The gene was cloned into pET29b (Novagen), protein L4THD7 was expressed in *E. coli* BL21 (DE3) and purified from the soluble cell extract via affinity chromatography on Ni<sup>2+</sup>-agarose (Qiagen). The histidine tag was cleaved off with thrombin and the resulting L4THD7 protein was further purified via ion exchange chromatography where it eluted from a Mono S column (Amersham Biosciences) at 270 mM KCl. The protein was via ion exchange chromatography where it eluted from a Mono S column (Amersham Biosciences) at 270 mM KCl. The protein was.

In Vivo Regulatory Studies—The regulatory capacity of the *T. aquaticus* L4 proteins was determined as described (7). A C-terminal 6-histidine tag (no thrombin cleavage site) was introduced into both L4 (wt) and L4 proteins was determined as described (7). A C-terminal 6-histidine tag (no thrombin cleavage site) was introduced into both L4 (wt) and L4-mediated transcription control (19). Plasmid pACYC-HD-S10 lacZ (Fig. 1A, red brackets). We also made a construct encoding a transcript with a hammerhead generating the C125 end.

**Nitrocellulose Filter Binding Assays—**Nitrocellulose filtration assays, described in Ref. 20, were done in 300-mM reactions containing 20 mM Hepes-KOH (pH 7.5), 4 mM MgCl<sub>2</sub>, 4 mM β-mercaptoethanol, 200–400 mM NH₄Cl, RNase-free BSA and tRNA bulk (as indicated). Binding data were fitted to single site hyperbolic curves (Fig. 1, upper panel). In competition assays, the L4 concentration was 0.67 pmol of complex was collected on the nitrocellulose filter.

**Gel Shift Assays—**A 15-µl sample of an RNA-protein complex containing 75 pmol of [<sup>35</sup>P]RNA and the indicated amount of protein L4 was prepared by incubating for 15 min at 37 °C (20 mM Hepes-KOH (pH 7.4), 100 mM NH₄Cl, 4 mM MgCl<sub>2</sub> 10% glycerol) and then chilling on ice. The sample was loaded onto a buffered (20 mM Hepes-KOH (pH 7.4), 100 mM NH₄Cl, 4 mM MgCl<sub>2</sub>) 5% acrylamide/bisacrylamide (29/1) gel (0.75 × 130 × 170 mm). The gel was run at 110 V for 2 h at 4 °C and the electrode buffer was recirculated between the electrode chambers with.

**Fig. 1. A, secondary structure of the *E. coli* S10 leader mRNA (6, 7) and constructs derived from it. Green arrows, with the additional Gs for efficient T7 RNA polymerase run-off transcription, indicate transcription start sites. Red brackets indicate the ends of the RNA constructs. Deviations from the wild-type *E. coli* sequence are indicated by red nucleotides. The six helical structures (HA, HB, HC, HD, IH, lower part of HE, upHE (upper part of HE), and HG), the Rho-independent transcription termination site (TR-termination), the Shine-Dalgarno sequence (SD), the translation initiation codon (TL-start), and the helix nomenclature are shown in blue. B, sequence alignment of *E. coli* L4 and L4 from *T. aquaticus*. Conserved amino acids are indicated in blue, residues putatively involved in RNA binding are shown in red (see text). The 46-amino acid deletion of protein L4THD7 within the extended loop is indicated in red (cf. Fig. 4C).
Interaction of Ribosomal Protein L4 with the S10 mRNA Leader

**RESULTS**

**Binding of Protein L4EC to the S10 mRNA Leader RNA in Vitro**—To analyze L4 interaction with the S10 leader, we modified the strategy previously used to demonstrate L4 binding to its target in Domain I of 23 S rRNA. Ribosomal protein L4EC was incubated with 32P-labeled S10 leader RNA, or, as a control, its known target in 23 S rRNA, and subjected to nitrocellulose filtration. The results in Fig. 2 (top panel) demonstrated that the full-length leader mRNA (A–G, 1–197C) bound r-protein L4 with an apparent $K_D$ of about 700 nM (in 4 mM MgCl$_2$, 250 mM NH$_4$Cl; binding was strongly dependent on the ionic condition). The affinity is in the same range as reported for the complex of ribosomal RNA with protein L4 (15). Phosphorothioate RNA was produced by including $S_{-\text{inc}}$-SINTPs (PerkinElmer Life Sciences and Glen Research) in the T7 RNA polymerase transcription assay. The L4 concentration was 0.67 $\mu$M and that of 32P-labeled phosphorothioated RNA was 0.53 $\mu$M or vice versa (the yield of complex was identical in both experiments). As seen in some experiments, the mRNA fragments tend to break at C38-A39 and C68-A64 even in the absence of iodine, and at G80-A81 in the presence of L4THD7. However, in general control experiments carried out without addition of iodine did not result in cleavages.

Because many highly positively charged ribosomal proteins, such as L4, bond with considerable strength to nonspecific RNAs (e.g. Refs. 25 and 26), we tested the specificity of L4 binding to the S10 leader by competition experiments (Fig. 2, middle panel). As expected, non-radioactive S10 leader RNA (A–G) competes with binding of radioactive leader RNA. In contrast, addition of a 218-nt mRNA molecule that contains the target in Domain II of 23 S rRNA (rRNA-DII) did not bind, even though this part of the 23 S rRNA contains a site that cross-links to L4 in the intact ribosome (24). Other control RNAs, e.g. fragments of ribosomal RNA of various lengths (50 to 170 nt), also failed to bind to r-protein L4 (data not shown). Likewise, the sequence comprising the first three helices from the beginning of the S10 leader sequence did not bind (Fig. 2, top, A–C, 1–66). This result suggests that the S10 leader contains one or more sites for L4 binding downstream of the HC hairpin.

To define the binding site more precisely, we tested various S10 leader mRNA subdomains in the competition experiments (Fig. 2, bottom panel). Most intriguingly, the mRNA fragments that do not contain helix HD, such as A–C (66 nt), E (65 nt), and E–G (116 nt), do not inhibit complex formation, whereas the isolated helix HD (D, 27 nt) competes significantly. All other constructs containing HD (viz. A–D (86 nt), D–E (88 nt), and D–G (139 nt)) compete to the same extent as the full-length A–G (198 nt). Consistent with data indicating the importance of helix HD in L4-dependent transcription regulation (11, 28), these experiments suggest that helix HD contributes the most to the specificity of L4 binding.

**T. aquaticus Protein L4 Actively Regulates E. coli S10 Operon Expression**—Protein L4 from E. coli is not well suited for in vitro binding experiments, because (i) it has to be purified...
under denaturing conditions either from ribosomes or from inclusion bodies, (ii) it is relatively unstable, (iii) its solubility even in high salt buffers is very low (<50 μM), and (iv) it aggregates even at low concentrations. For example, in gel-shift experiments RNA-L4EC complexes form aggregates that remain for the most part in the gel pocket (not shown).

To circumvent these problems we combined two strategies to develop a better behaved protein. First, because L4 proteins from many other eubacteria regulate the E. coli S10 operon (7, 19), we switched to L4 from T. aquaticus (L4TH), which is 34% identical and 50% similar to L4 from E. coli (19). Proteins from such thermophilic bacterial species have proven to be beneficial for other biochemical and biophysical experiments. Second, we deleted 46 amino acids that, in the ribosome, form an extension protruding into the core from a globular domain located on the surface of the ribosome (29, 30) (cf. Figs. 1B and 4C). In free L4 protein (from Thermotoga maritima) the corresponding amino acids are not organized into a particular structure (31), and hence might contribute to the aggregation problem. Indeed, the deletion of this region in E. coli L4 generates a protein that retains its capacity for regulating the S10 operon, but is more soluble than the full-length protein. Therefore, we constructed similar deletion derivatives of the T. aquaticus L4 and screened them for solubility and thermostability. A deletion between Gly45 and Pro92 of L4TH (corresponding to Gly42-Pro69 in L4EC) rendered the protein (called L4THD7) very soluble and stable.

Before using L4THD7 for S10 mRNA binding studies we confirmed that the protein is active in inducing autogenous control in E. coli. First, E. coli cells carrying plasmids with either L4TH or L4THD7 under control of an arabinose-inducible promoter could not form colonies on plates containing arabinose, presumably because the induced T. aquaticus L4 proteins reduce expression of the 11 r-protein genes in the E. coli S10 operon, thereby blocking synthesis of ribosomes.

Second, we deleted 46 amino acids that, in the ribosome, form a protein that retains its capacity for regulating the S10 operon and has the desired solubility characteristics. Highly purified protein L4THD7 was therefore used in vitro studies.

**Table I**

In vivo analysis of autogenous regulation by T. aquaticus L4 in E. coli

| L4 protein | L4 regulation of S10L-S10′/lacZ | L4 regulation of S10L-S10′/lacZ |
|------------|---------------------------------|---------------------------------|
| L4EC       | 0.09 (±0.02)                    | 0.54 (±0.02)                    |
| L4TH       | 0.31 (±0.03)                    | 0.71 (±0.06)                    |
| L4THD7     | 0.41 (±0.07)                    | 0.81 (±0.06)                    |

The average is shown, with the standard deviation given in parentheses.

 analogue of the E. coli L4 with respect to repression of the E. coli S10 operon and has the desired solubility characteristics. Highly purified protein L4THD7 was therefore used in vitro studies.

**RNA Determinants for L4 Binding**—To determine a minimal binding site in the S10 leader mRNA, RNA leader fragments were assayed for binding to L4THD7 in gel shift assays (Fig. 3). S10 leader mRNA fragment D (GG62–86) does not bind to protein L4 in the gel shift experiment, and there is almost no binding to the D-IoE fragment (GG64–100). The presence of helix E appears to be more important in the band shift assays than apparent from the nitrocellulose filtration experiments (cf. Fig. 2, bottom), because a little binding activity is seen with the complete helix E stem-loop structure (E, GG86–147). Nevertheless, the minimal mRNA that shows high affinity binding is D-upE (64 nt), comprising nucleotides GG64–125 (Fig. 3, top). The binding of the D-upE part to L4THD7 deter-
mined by nitrocellulose filter binding is shown in Fig. 1 for comparison (upper panel, black squares). Binding is about 2-fold weaker than binding of L4EC to the full-length mRNA fragment.

Together with HD, the upper helix HE (nt 100–125) is strictly required for high affinity L4 interaction. The D-upE mRNA fragment, which shortcuts helix HE by connecting C100 and G132 with a UUGC-tetraloop (Fig. 1), does not bind (cf. D–E and D-upE in Fig. 3, middle). The primary sequence requirements of the upper HE are not strict, because replacement of the upHE-loop with an U1A-loop (Fig. 1) retains significant affinity for L4 (D-U1AupE in Fig. 3, middle).

The gel shift experiments with the L4THD7 are in good agreement with the nitrocellulose filtrations where full-length L4EC has been used, although the apparent dissociation constants are higher. However, gel shift assays are known to generate apparent K_D values up to 1 order of magnitude higher than filter binding assays (32). Similar to the nitrocellulose competition experiments presented above, A–D and D–G bind to L4THD7 in the gel shift system (not shown). Furthermore, binding is stronger with the cognate ribosomal binding site (RNA-DI) and no binding can be detected with the dispensable A–C part of the S10 leader mRNA (Fig. 3, bottom). In summary, a minimal 64-nt fragment comprising HD, the ascending side of HE, and the upper HE helix is sufficient for high affinity binding of protein L4.

L4-S10 Leader mRNA Contacts Monitored by Iodine Footprinting—Iodine cleavage of phosphorothioated RNA is a useful tool for examining in detail RNA-protein interactions (15, 33, 34). To execute such experiments one α-phosphorothioate nucleotide (A, C, G, or U) per 5–32P-labeled transcript molecule, on average, was introduced to RNA via T7 in vitro transcription. The phosphorothioate S10 leader transcripts were used for two types of experiments. In an interference experiment, RNA-protein complexes were formed with phosphorothioate-containing RNA and isolated by binding to nitrocellulose filters. The RNA was then extracted, cleaved with iodine, and subjected to gel electrophoresis. A band that is weakened or missing on the gel indicates that the presence of a phosphorothioate moiety at the corresponding position impairs or prevents the formation of the RNA-protein complex. In a protection experiment, the RNA is cleaved in the complex preceding its isolation. A weakened or absent band in this experiment indicates that iodine does not have free access to the corresponding phosphorothioate in the backbone of the RNA, i.e. the position is protected from iodine cleavage. In this way protein interactions with the non-bridging Rp-oxygens of the RNA phosphate backbone can be mapped with atomic resolution.

Full-length S10 leader A–G and the two S10 leader fragments A–D and D–G were probed in complex with protein L4EC. Additionally, protein L4THD7 has been footprinted in complex with the minimal D-upE mRNA fragment, resulting in a pattern similar to that obtained with protein L4EC on the larger fragments.

Fig. 4A shows a section of the iodine cleavage pattern of the D–G region that contains all determinants necessary for transcription and translation control by L4. Taken together, all signals are located in helices HD and HE clearly monitoring the L4 binding site. Protein L4EC shows strong interference with 5′-phosphorothiolated U75, C76, C86, and to C94 (Fig. 4B, stars), and weaker interference at the positions C72, A73, G78, and A99. These same regions are also protected. Non-bridging Rp-oxygens 5′ to the following nucleotides are protected against iodine cleavage in the complex with L4EC: U66, G67 (note: not shown in Fig. 4A), U69, and A74 in the 5′-strand of helix HD; in A81/85, C86-U88, A90-G95, and U97-A99 starting from HD extending to the 5′ strand of lower helix HE and, finally, C111, C115, U117, U124, and G132 (not shown) in the upper helix HE (Fig. 4B, dots). The signals cover almost every phosphate in the backbone of the 5′ part of lHEU up to C100.

This extensive protection could either be due to direct protein contact in this region or to structural changes induced by L4 binding to the HD region (see “Discussion”). However, the footprint in the loop of helix HD displays very characteristic protein-RNA contacts. Interestingly, the two interference sites 5′ and 3′ to U75 (Fig. 4B, red stars) are strongly reminiscent of the footprinting pattern of rRNA-DI in complex with L4, where O to S substitutions 5′ and 3′ to U321 abolish the binding of L4 (14) (Fig. 4B, inset). The three phosphates 5′ to A74 and A320 and G78 and A324 are protected in the mRNA and the rRNA, respectively (Fig. 4B, red dots). Although the secondary structures are different, these data raise the intriguing possibility that helix HD of the S10 leader is a close three-dimensional mimic of the rRNA binding site (U321-loop, Fig. 4B).

Mutational Analysis of the HD Loop—Because the footprinting data suggested direct L4 binding to the loop of hairpin HD, and because the pattern was strongly reminiscent of the interaction of L4 with its rRNA target, we analyzed the sequence requirements for L4 interaction with hairpin HD in more detail. Comparison of the E. coli HD loop structure with the homologous structure in S10 leader mRNAs from other γ-proteobacteria (7), and with the U321 loop in ribosomal RNA (L4-binding site) of different species reveals the following (Fig. 5). (i) A74 in the E. coli mRNA, which corresponds to position 320 in rRNA, is conserved in the S10 leader of other proteobacteria that bestow L4-mediated autogenous control; (ii) the rRNA bases 319 and 323 universally form a Watson-Crick base pair (e.g. G–C in E. coli and Haloarcula marismortui, and C–G in T. thermophilus and Deinococcus radiodurans), whereas the corresponding bases at positions 73 and 77 in the mRNA occur as noncanonical A–G or G–U combinations; and (iii) the nucleotide corresponding to C76 in E. coli mRNA is an adenine in 23 S rRNA throughout all three kingdoms of life (A322 in E. coli). Guided by these comparisons (Fig. 5) and assuming that the two L4 binding sites form similar structures, we constructed 14 mutant mRNA minimal versions and tested them for binding to protein L4THD7 in the band shift system. Selected experiments are shown in Fig. 5, together with a quantitative representation of the binding experiments in Table II.

The binding data support the idea that L4 recognizes specific primary or secondary structure features in the HD loop region. First, substitutions of the conserved A at position 74 decrease the binding affinity (Table II, mutants M1–3), particularly if the base is changed to U (Fig. 5). Second, we were interested in nucleotides 73 and 77 in the HD loop, because they are Watson-Crick paired in the 23 S rRNA target. Interestingly, the affinity for L4 for the leader was significantly reduced when bases in those positions were changed to Watson-Crick combinations (Table II, mutants M5–8), but additional changes such as C76A (M9 which resembles E. coli rRNA) or changes weakening the loop-closing C72–G78 base pair (M10–14) partially or fully restored binding affinity. The C76A mutation assayed by itself increased binding slightly above wild type (M4).

In summary, base exchanges in the loop of helix HD modulate the affinity of the S10 leader mRNA to L4 as much as 3–4-fold. Consistent with in vivo data suggesting that A74 is important, and that requirements for the HD sequence are flexible (28), none of the changes completely eliminates L4 binding. The mutational analysis reflects the phylogenetic conservation of the two RNA sites.
DISCUSSION

Regulatory Features of Protein L4 Are Conserved in T. aquaticus—The model for autogenous control of the expression of an r-protein operon assumes dual RNA binding activity of the regulating r-protein (13). That is, the protein has binding sites with affinity for both the mRNA and rRNA (1, 2, 35). Regulation of the S10 operon transcription has been analyzed in vivo and in an in vitro transcription termination system, but proof for the presumed interaction between the S10 leader and the L4 protein has remained illusive. The high propensity of L4 for aggregation and unspecific affinity to nucleic acids have in the past thwarted attempts to demonstrate a direct interaction between L4 and leader mRNA (1, 16).

The current investigation therefore focuses on analyzing L4 interaction with the S10 leader. To circumvent the difficulties discussed above we used competition strategies and developed an L4 protein to overcome aggregation and precipitation problems. The construction of this protein, derived from T. aquaticus, took advantage of the superior stability of r-proteins from thermophiles, and the fact that the long loop covering about 50 amino acids in L4 (cf. Fig. 4C) (29, 30) apparently contributes to the undesirable properties of the protein, but is dispensable for regulation (17). The resulting protein, L4THD7, with a 46-amino acid deletion in the extended loop, facilitated biochemical experiments.

The ability of T. aquaticus L4 to regulate the E. coli S10 operon is interesting, because the T. aquaticus S10 operon is fused with the upstream str operon, and hence has no S10 leader (36). One must therefore assume that L4TH does not regulate the expression of the S10 operon in T. aquaticus, at least not by the mechanisms described for E. coli, yet the protein has maintained the determinants for that function. Similar findings have previously been made for L4 proteins from bacterial species such as Pseudomonas aeruginosa and...
Bacillus stearothermophilus, species that also lack an E. coli-like S10 leader (7, 19). Taken together, these findings suggest that determinants necessary for the interaction of L4 with the S10 operon mRNA of E. coli are conserved in other eubacteria, we presume because these same determinants are involved in ribosome assembly and/or function in the native background. This conclusion is consistent with our hypothesis that the targets of L4 in the S10 mRNA and in 23 S rRNA share structural features.

The S10 Leader mRNA Binding Site of Protein L4—In this study we demonstrated for the first time direct and independent binding of L4 to the S10 leader RNA, using both nitrocellulose filter competition binding assays (Fig. 2) and gel shift experiments (Fig. 3). The minimal binding site for L4 binding consists of a 64-nt fragment containing hairpin HD, the 5′ part of the lower HE, and the stem loop of the upper HE (Fig. 3, top). We propose that L4 binding to the leader requires specific interactions with the HD loop and non-sequence specific interaction with parts of HE.

The requirement for HD is supported by the following results: (i) no leader fragment lacking hairpin HD has significant binding or competition activity (Figs. 2 and 3); (ii) HD competes by itself against binding to other, longer leader fragments and thus seems to contribute the most to the specificity of binding (Fig. 2, bottom); and (iii) point mutations in the loop of hairpin HD positively and negatively alter binding affinity (Fig. 5, Table II). Thus, hairpin HD is essential for binding and seems to be the main determinant for specificity. This result is consistent with in vivo genetic experiments showing that L4 transcription control is abolished upon deletion of hairpin HD (Table I and Ref. 11).

Nevertheless, HD does not bind by itself, so other sequences must contribute to binding. Because both A–D and D–G bind (Fig. 2), this extra energy can apparently be provided by sequences on either side of HD. The relatively weak effect of substituting the loop of HE by the U1A-loop (D-U1AupE, a change of 11 bases) implies that these sequences can add binding energy but do not contribute significantly to specificity. Deletion of the upper HE part, on the other hand, abolishes binding (D-upE, Fig. 3, middle). This requirement might be because of indirect effects, e.g. improper folding of the actual L4 target, as the upper HE is not protected by L4 binding. Consistently, other hairpins with an intact stem can substitute for the native upper HE hairpin with little or no loss of activity in vivo (8, 28). Nonspecific interactions have also been observed to be an essential part of specific binding and function in other RNA-protein complexes (e.g. Ref. 37).

Structural Mimicry between mRNA and rRNA—The phosphorothioate footprinting results further define the hairpins HD and HE as the binding target of r-protein L4 (Fig. 4). Surprisingly, almost every phosphate in the backbone of the 5′ part of loHE up to C100 is protected. We cannot differentiate whether the backbone is directly shielded from iodine cleavage by protein L4 or the footprint indicates changes in the RNA structure. We favor the latter possibility because it is difficult to see how the protein could protect one strand of the lower helix HE for more than one turn. Rather, the results suggest that in complex with L4, even in fully transcribed RNA and in the absence of an RNA polymerase complex, lower helix HE as such does not exist (cf. Fig. 1A versus 4B), and that its structure is influenced by L4 binding. Notably, secondary structure probing results obtained with the free RNA have been ambiguous for this particular region (6).

On the other hand, the footprint in the loop of helix HD is

**Fig. 5.** Phylogenetic comparison of the HD-loop in S10 mRNA leader sequences and the rRNA U321-loop of Domain I of 23 S rRNA. mRNA-HD1 is found in E. coli (C76), Salmonella typhimurium (C76), and Citrobacter freundii (U76). mRNA-HD2 occurs in Yersinia enterocolitica (A76), Serratia marcescens (A76), Morganella morganii (A76), and Haemophilus influenzae (U76). All these leader sequences are conserved in other eubacteria, Boxed, gel shift assays (5% polyacrylamide gel, 20 mCi Hepes-KOH (pH 7.4), 100 mM NH4Cl, 4 mM MgCl2, 0.01 mg/ml AcBSA). For every RNA mutant (3 lanes each), protein L4THD7 was added to 7.5, 12, and 17 mM Hepes-KOH (pH 7.4), 100 mM NH4Cl, 4 mM MgCl2, 0.01 mg/ml AcBSA). For every RNA mutant (3 lanes each), protein L4THD7 was added to 7.5, 12, and 17 μM. [53P]RNA mutants (5 μm) are D-upE minimal fragments. Mutated nucleotides are indicated on the black background in the HD-loop secondary structures (see also Table II).

**Table II**

Quantitative analysis of the effect of mutations in the loop of helix HD on the interaction with LATHD7

| Name/position | 72 | 73 | 74 | 75 | 76 | 77 | 78 | Binding +/− |
|---------------|----|----|----|----|----|----|----|----------|
| D-upE-wt      | C  | A  | U  | C  | G  | G  | 1.00|          |
| M1            | U  |    |    | C  |    |    | 0.48| 0.10     |
| M2            | C  |    |    |    |    |    | 0.66| 0.15     |
| M3            | G  |    |    |    |    |    | 0.76| 0.15     |
| M4            |    | A  |    |    |    |    | 1.12| 0.15     |
| M5            | C  |    |    |    |    |    | 0.31| 0.07     |
| M6            | G  |    |    | C  |    |    | 0.49| 0.18     |
| M7            |    | U  |    |    |    |    | 0.56| 0.13     |
| M8            | U  |    |    | A  |    |    | 0.28| 0.09     |
| M9            | G  |    |    | A  |    |    | 0.84| 0.10     |
| M10           | A  |    |    | C  |    |    | 0.95| 0.09     |
| M11           | G  |    |    | C  |    |    | 1.07| 0.06     |
| M12           | A  |    |    | C  |    |    | 0.68| 0.19     |
| M13           | G  |    |    | C  |    |    | 0.63| 0.17     |
| M14           | U  |    |    | A  |    |    | 0.68| 0.22     |
particularly striking because the pattern is similar to the signals that characterize the interaction of L4 with the U321-loop of 23 S rRNA (14). The rRNA binding site of L4 contains a four-helix junction (rRNA-DI, Fig. 4B), which, in the ribosome (29, 30, 38), forms two longer bent helices: helices H118 and H20 stack on each other generating one continuous helix, whereas H19 stacks on the U321-loop to form the other (Fig. 4C). This way three non-consecutive RNA strands (color coded in Fig. 4, B and C) contribute bases to the helix-like arrangement that is recognized by L4: U304–G310 (green) and C334–G333 from H20 (red) base pair with G313–C318 (blue). Stacking on top of this helix stem, the G319–C323 pair closes the three-base U321-loop (cyan). Our interference results raise the interesting possibility that helix HD of the S10 leader RNA has a three-dimensional structure similar to the structure of the H19–U321-loop part of the four-helix junction: the two interference sites 5’ and 3’ to U75 resemble the phosphates 5’ and 3’ to U321 in 23 S rRNA, where O to S substitutions abolish L4 binding, and the preceding three phosphates are protected, as is the phosphate 3’ to U75 and U321 in mRNA and rRNA, respectively (Fig. 4, red dots and stars). The hypothesis that hairpin HD is a three-dimensional structure mimic of the rRNA-DI binding site is supported by our observation that binding of mRNA and rRNA are competitive.

The Watson-Crick 319–323 pair closing the three-base U321-loop in rRNA (Fig. 4D) is phylogenetically conserved in all three domains of life (Fig. 5), whereas the phylogenetic analysis of the mRNA suggests that the HD loop has five or more nucleotides (7). In an attempt to change the mRNA structure to a structure more like the rRNA, we changed nt A73 and/or G77 to form Watson-Crick base pairs. Most pronounced in the pyrimidine-purine combinations (i.e. M5 and M8), mutations predicted to form 73–77 Watson-Crick base pairs in the HD loop negatively affect the interaction with L4 (Fig. 5, Table II, M5–M8). In the crystal structures of the 50 S ribosomal subunit, the 3’ nucleotide of the Watson-Crick pair (C330 of \textit{H. marismortui} (38) and G334 of \textit{D. radiodurans} (30)) is in a syn-conformation causing a turn between this and the preceding base (Fig. 4D). It appears difficult to form a Watson-Crick base pair with one base in the syn-conformation in the context of an A-form helix like HD, because it would similarly require a local reversal of strand direction (39).

Whereas keeping the G73–G77 Watson-Crick pair, we tried to provide a context to form a three-base loop with a similar conformation as in the rRNA by weakening and destroying the last base pair of the helix HD. Weakening of the closing base pair in the stem of HD, i.e. substitution of C72–G78 with A–U or U–A, indeed increased the binding affinity (M12 and M13). Replacing the C72–G78 base pair by a non-Watson-Crick combination even restores binding of the G73–C77 mutant version to wild-type levels (M10, M11). Thus, the particular backbone conformation of the three-base loop might be an important determinant for recognition of L4. One might speculate that the HD-loop does indeed only contain three nucleotides because of formation of A–G or G–U base pairs between positions 73 and 77 (Fig. 5) that could, in contrast to a Watson-Crick combination, enable a similar shape of the three-base loop.

Although a three-dimensional structure of the mRNA-L4 complex is clearly needed to substantiate these speculations, it is interesting to note that the backbone turn between A322 and C323, caused by the syn-conformation of the 323 base in the rRNA, is monitored in the probing results. Because of this turn, the phosphate group 5’ to C323 is facing opposite to the L4 side, and is indeed the only phosphate in the loop hardly affected in the probing experiments of both the mRNA and rRNA L4 complexes (Fig. 4D, green).

\textbf{Analogies between the L4 Contacts to mRNA and rRNA Targets}—Although L4 interacts with parts of Domains I, II, IV, and V in the mature 50 S subunit (30, 38), the only part of 23 S rRNA that binds to L4 in the absence of other proteins is Domain I (15). Proposing that the two RNA targets for L4 binding have a similar structure, one would expect the same protein L4 to interact with each of the RNA targets.

Phylogenetically conserved amino acids in the α4/α5 regions of L4 (Fig. 1B) contact the RNA backbone in 23 S rRNA Domain I. In the \textit{H. marismortui} ribosome (38) (Protein Data Bank number 1JJ1), residues Lysα4, Leuα5, Alaα6, and Valα5 abolish L4-dependent regulation in \textit{in vivo} (17). Furthermore, modifications in the α4/α5 region of L4 from \textit{T. maritima} bestow on this protein a capacity to regulate the \textit{E. coli} S10 operon not possessed by wild-type \textit{T. maritima} L4 protein (40). Thus, amino acids in L4 that contact 23 S rRNA Domain I or are close to these L4 residues may also be involved in recognizing the HD mRNA region, supporting the analogy between the two binding events.

\textbf{Conclusions on the Mechanism of L4-mediated Transcription Termination}—Studies of Rho-independent transcription termination suggest that cessation or prolongation of transcription is caused by changes in the conformation of the nascent RNA chain (41, 42). These changes could be effected by protein factors, such as NusA and AN protein (43, 44), or ribosomal proteins (45). Analysis of transcription complexes paused at other terminators (46, 47) suggests that exactly the minimal D-upE segment would be accessible to L4 binding in the paused complex. L4-induced termination of transcription at the S10 attenuator requires both L4 and NusA (12, 49). In the current study we have found that L4 can bind to the S10 leader mRNA in the absence of NusA. The effect of NusA is thus not related to L4 binding. Our footprint data suggest changes in the RNA leader structure induced by L4 binding. If so, NusA might be required for propagation of this signal to the RNA polymerase during the transcription termination event.

\textbf{Acknowledgments}—We thank Beatriz Cebolla for help with initial experiments, Alexander Serganov for the pAB5313 S15-mRNA plasmid, and Alexander Mankin for the \textit{L. T. aquaticus} clone.

\textbf{REFERENCES}

1. Zengel, J. M., and Lindahl, L. (1994) \textit{Prog. Nuclear Acids Res. Mol. Biol.} \textbf{47}, 311–370.

2. Keener, J., and Nomura, M. (1996) in \textit{Escherichia coli and Salmonella typhimurium} (Neidhardt, F. C., ed) pp. 1417–1431, American Society for Microbiology, Washington, D. C.

3. Zengel, J. M., Mueckl, D., and Lindahl, L. (1980) \textit{Cell} \textbf{21}, 523–535.

4. Yates, J. L., and Nomura, M. (1980) \textit{Cell} \textbf{21}, 517–522.

5. Lindahl, L., Archer, R., and Zengel, J. M. (1983) \textit{Cell} \textbf{33}, 241–248.

6. Shen, P., Zengel, J. M., and Lindahl, L. (1988) \textit{Nucleic Acids Res.} \textbf{16}, 8905–8924.

7. Allen, T., Shen, P., Samuel, L., Liu, R., Lindahl, L., and Zengel, J. M. (1999) \textit{J. Bacteriol.} \textbf{181}, 6124–6132.

8. Freedman, L. P., Zengel, J. M., Archer, R. H., and Lindahl, L. (1987) \textit{Proc. Natl. Acad. Sci. U.S.A.} \textbf{84}, 6516–6520.

9. Zengel, J. M., and Lindahl, L. (1990) \textit{J. Mol. Biol.} \textbf{213}, 67–78.

10. Sha, Y., Lindahl, L., and Zengel, J. M. (1995) \textit{J. Mol. Biol.} \textbf{245}, 486–498.

11. Zengel, J. M., and Lindahl, L. (1996) \textit{J. Bacteriol.} \textbf{178}, 2385–2387.

12. Sha, Y., Lindahl, L., and Zengel, J. M. (1998) \textit{J. Mol. Biol.} \textbf{285}, 474–485.

13. Nomura, M., Yates, J. L., Dean, D., and Post, L. E. (1986) \textit{Proc. Natl. Acad. Sci. U.S.A.} \textbf{77}, 7084–7088.

14. Steihl, U., Spaht, C. M. T., and Nierhaus, K. H. (2000) \textit{Proc. Natl. Acad. Sci.}
RNA-structural Mimicry in *Escherichia coli* Ribosomal Protein L4-dependent Regulation of the *S10* Operon
Ulrich Stelzl, Janice M. Zengel, Marina Tovbina, Marquis Walker, Knud H. Nierhaus, Lasse Lindahl and Dinshaw J. Patel

*J. Biol. Chem.* 2003, 278:28237-28245.
doi: 10.1074/jbc.M302651200 originally published online May 8, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M302651200

**Alerts:**
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 48 references, 20 of which can be accessed free at http://www.jbc.org/content/278/30/28237.full.html#ref-list-1