*Molecular Biology, Vol. 39, No. 6, 2005, pp. 942–945. Translated from Molekulyarnaya Biologiya, Vol. 39, No. 6, 2005, pp. 1072–1075.*


**SHORT COMMUNICATIONS**

*UDC 577.053*

# Expression Regulation of the Amino Acid Biosynthesis and Aminoacyl-tRNA Synthase Genes in Actinobacteria

**V. A. Lyubetsky and A. V. Seliverstov**

*Institute of Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia*
*e-mail: lyubetsk@iitp.ru*

Received April 5, 2005

*Key words*: gene expression, attenuation, translational regulation, T-box, multiple sequence alignment

The formation of alternative RNA structures in response to external factors is an important mechanism regulating the expression of bacterial genes. Comparison of the 5'-leader gene regions revealed conserved RNA structures. Attenuating regulation was predicted for the tryptophan and cysteine biosynthesis operons. An element forming a conserved secondary structure in RNA was found upstream of *leu*A and termed LEU. Translational regulation involving the T-box was predicted for *ile*S.

The nucleotide sequences of bacterial genomes were extracted from GenBank (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). Multiple sequence alignments were constructed with regard to the RNA secondary structure. Computations involved published algorithms and programs [1, 2] along with the MultAlign program (A.A. Mironov). In the Clique program [2], word length, as isolated in each sequence, was approximately 12 nt for the LEU element. The upper threshold of the difference between two words was approximately 4 nt. Other values of these parameters were also considered, depending on the number of sequences expected to produce a signal. The MutlAlign program was run with default parameters: Length = 20, Mismatches = 9, min Score = 2.8, nlter = 3, Gap Param = 8.0, min Inf = 0.8, Top = 60%, min Energy = 6.5, max loop = 80, Nested Dist = 1, Signal length = 12, and Dist max = 8. The default parameters of the Spiral program [1] were Kolraz = 1, Limit = 10, MaxVypach = 5, MinPetlya = 3, MaxPetlya = 15, MinPodryad = 3, VNihGT = 3, MaxGT = 3, EnergPorog = 15, and DlinPorog = 5.

**Tryptophan.** Continuing our previous study [3], we observed classical attenuation of the tryptophan biosynthesis operons in all *Corynebacterium* and *Streptomyces* species. Classical attenuation was predicted for two *C. diphteriae* operons, $trpB_1EDGC$ and $trpB_2A$, and *S. avermitilis* $trpS_2$, coding for tryptophanyl-tRNA synthase.

**Cysteine.** It is known that the 5'-terminal regions of the *Mycobacterium* spp. and *Propionibacterium acnes cys* operons and the *Bifidobacterium longum cbs* operon contain open reading frames with cysteine codons immediately upstream of the stop codon. Transcriptional regulation of these operons is probably based on Rho-dependent termination, as is the case with the *Escherichia coli tna* gene coding for tryptophanase [5, 6]. The first genes of the *M. avium* and *M. laprae* cysteine biosynthesis operons are hypothetical (*MAP2122* and *ML0840*, respectively) and code for orthologous proteins.

**Leucine biosynthesis.** The 5'-terminal region of *leu*A, coding for isopropylmalate synthase, contains a characteristic conserved structure in most actinobacteria (*Actinomyces naeslundii, Corynebacterium* spp., *Kineococcus radiotolerans, Leifsonia xyli, Mycobacterium* spp., *Nocardia farcinica, Streptomyces* spp., and *Thermobifida fusca*). The structure was termed LEU element (Fig. 1). The LEU element includes a short open reading frame, which codes for a leader peptide and contains several Leu codons, and the downstream nucleotide sequence. In mRNA, the LEU element can form two alternative secondary structures. One is a secondary structure with a pseudoknot that has a highly conserved nucleotide composition of the arms of its helices; the other does not form a pseudoknot (Fig. 2). The alternative structures are within the loop of the helix the 5' arm of which covers the region of the leucine codons of the leader peptide and the 3' arm covers the Shine–Dalgarno sequence of *leu*A. The following hypothesis can be assumed for the role of the LEU element in regulating gene expression. The stem of the hairpin is stable in the pseudoknot structure. In the alternative structure, the stem is unstable and the Shine–Dalgarno sequence is open. The LEU element was also found in the *B. longum* open reading frame coding for hypothetical transposase. Conserved nucleotides are few in the LUE element, although the structural stability of this element can be expected to selectively depend on the concentration of leucine but not of isoleucine and valine. The regulation of gene expression possibly

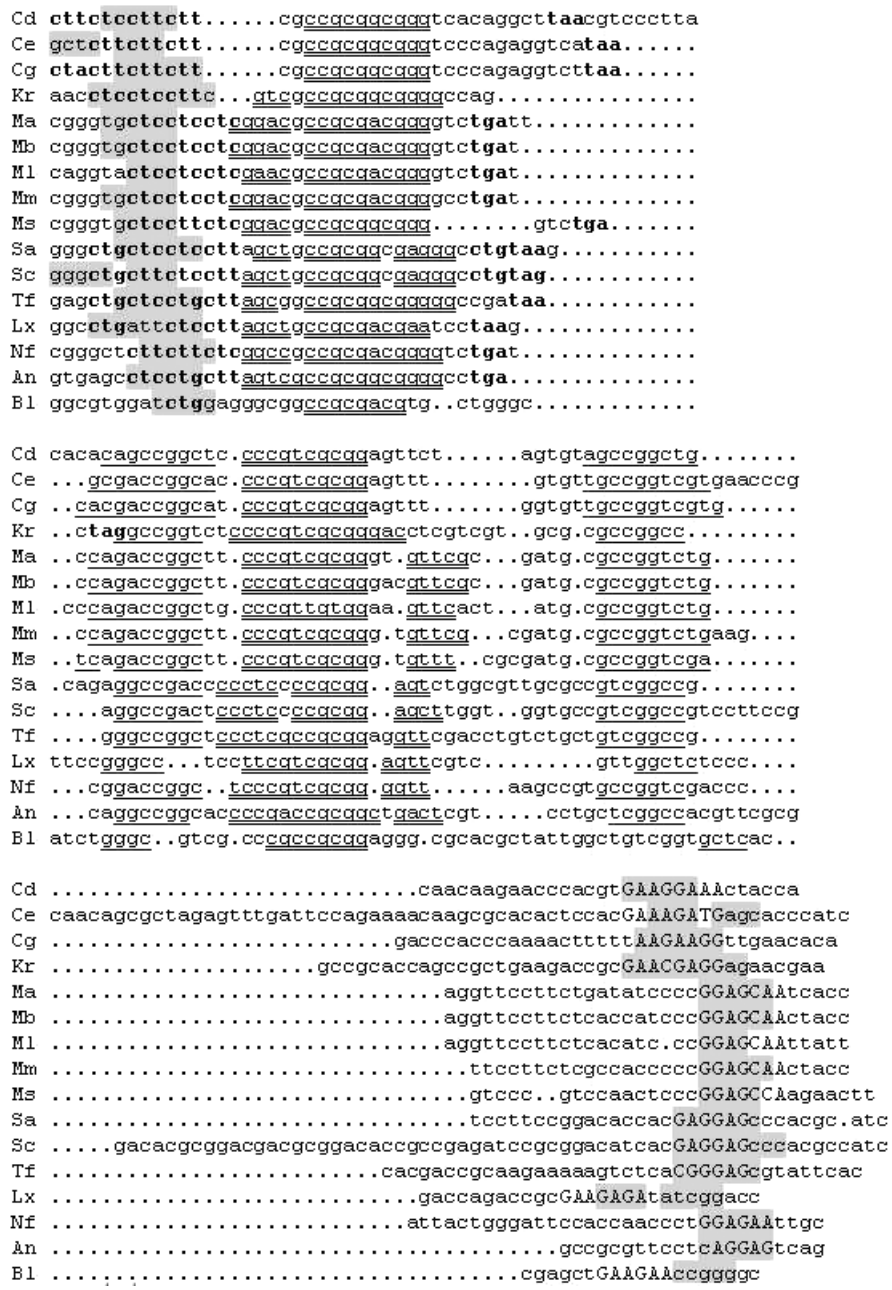

**Fig. 1.** Structure of the LEU element in the case of pseudoknot formation. Helices of the pseudoknot are underlined, stem nucleotides are shadowed, the Shine–Dalgarno sequence is in capitals, and cysteine codons and stop codons of the putative leader peptide-coding sequence are boldfaced.

involves a protein that complexates with the leucine molecule and thereby contributes to the formation of the pseudoknot by the LEU element. The phylogenetic profile closest to that of the LEU element is characteristic of homologs of the hypothetical *M. leprae* protein ML1624 (596 amino acid residues). Homologs of this protein with a E-value of less than $10^{-170}$ were found in all actinobacteria that have the LEU element upstream of *leu*A. All other bacteria lack close homologs of ML1624. Only *P. acnes*, which contains no LEU element in the 5'-leader region of *leu*A, has a homolog with a E-value of about $10^{-108}$, but homology is considerably lower in this case. Analysis with the PFAM database showed that ML1624 has the N-terminal domain (residues 34–193) characteristic of the DEAD/DEAH box of helicases. This domain is present in many proteins involved in RNA metabolism. It is possible to assume that, in response to a high leucine concentration, ML1624 expedites the pseudoknot formation in the LEU element of the *leu*A

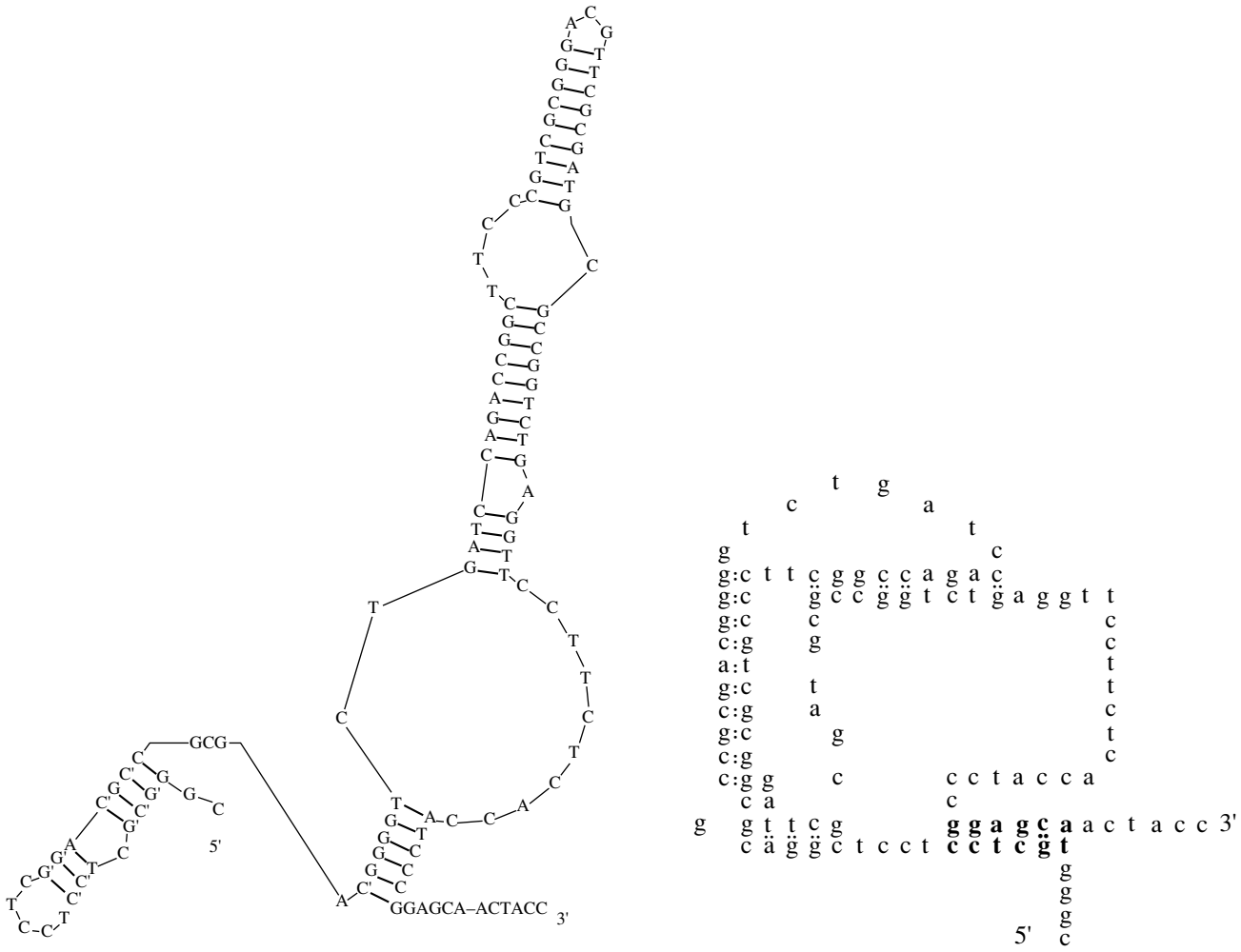


**Fig. 2.** Alternative structures of the *M. bovis* LEU element without (capital letters) and with (small letters) a pseudoknot. The Shine–Dalgarno sequence GGAGCA is open in the pseodoknot-lacking structure and closed in the alternative structure with the pseudoknot.

transcript, which sequesters the Shine–Dalgarno sequence and prevents translation. In all but one case (*B. longum leu*A), the 5'-terminal region of the LEU element contains an open reading frame with a region harboring leucine codons. The regulatory role of these codons is unclear. Their mutations exert no apparent effect on *leu*A expression in *S. coelicolor* [7]; however, it is in this bacterium that the stem of the LEU element is displaced and involves adjacent nucleotides, which are beyond the leucine codons. Another interesting feature of the LEU element is its presence in the *B. longum* transposase gene. This can be explained by multiple horizontal transfer of the LEU element contained in a transposon. The transposase gene is preserved in *B. longum* and has been transformed into a regulatory element of *leu*A in other actinobacteria during evolution.

**Aminoacyl-tRNA synthases.** Translational regulation involving T-box was predicted for *ile*S, coding for isoleucyl-tRNA synthase, in most actinobacteria (*A. naeslundii, B. longum, Corynebacterium* spp., *K. radiotolerans, Mycobacterium* spp., *N. farcinica, P. acnes, Rubrobacter xylanophilus*, and *Streptomyces* spp.). Regulation of this type does not involve a terminator, in contrast to standard transcriptional regulation [8, 9]. Free tRNA stabilizes the mRNA secondary structure with the Shine–Dalgarno sequence open for translation initiation. In the alternative structure, the ribosome-binding site is contained in an extended helix, which prevents translation of the *ile*S mRNA. The T-box has been identified earlier in *S. coelicolor* (H. Putzer, personal communication, which was received after the above T-boxes had independently been revealed by the authors). Potential canonical attenuating structures with a leader peptide-coding sequence, a terminator, and an antiterminator were found upstream of *leu*S (leucyl-tRNA synthase) in *S. avermitilis* and *S. coelicolor* and upstream of *trp*$S_2$ (tryptophanyl-tRNA synthase) in *S. avermitilis*. Tran-

scriptional regulation with the standard T-box was predicted for the *ala*S, *ile*S, *leu*S, *lys*$S_1$, *trp*S, and *tyr*S aminoacyl-tRNA synthase genes of *Symbiobacterium thermophilum.*

## ACKNOWLEDGMENTS

We are grateful to M.S. Gelfand for his guidance in this work and to K.Yu. Gorbunov for his participation in helpful discussion.

## REFERENCES

1. Gorbunov K.Yu., Mironov A.A., Lyubetsky V.A. 2003. Search for conserved secondary structures of RNA. *Mol. Biol*. **37**, 850–860.
2. Lyubetsky V.A., Seliverstov A.V. 2003. Some algorithms related to terminal groups. *Inform. Protsessy.* **3**, 39–46.
3. Vitreschak A.G., Lyubetskaya E.V., Shirshin M.A., Gelfand M.S., Lyubetsky V.A. 2004. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: Comparative genomics analysis. *FEMS Microbiol. Lett*. **234**, 357–370.
4. Seliverstov A.V., Lyubetsky V.A. 2004. Specific features of cysteine synthesis in *Corynebacterium, Mycobacterium, and Propionibacterium. Inform. Protsessy.* **4**, 247–250.
5. Konan K.V., Yanofsky C. 2000. Rho-dependent transcription termination in the *tna* iperon of *Escherichia coli*: Roles of the boxA sequence and the *rut* site. *J. Bacteriol*. **182**, 3981–3988.
6. Gong F., Yanofsky C. 2003. Rho's role in transcription attenuation in the *tna* operon of *E. coli. Methods Enzymol*. **371**, 383–391.
7. Craster H.L., Potter C.A., Baumberg S. 1999. End-product control of branched-chain amino acid biosynthesis genes in *Streptomyces coelicolor* A3 (2): Paradoxical relationships between DNA sequence and regulatory phenotype. *Microbiology.* **145**, 2375–2384.
8. Panina E.M., Vitreschak A.G., Mironov A.A., Gelfand M.S. 2001. Regulation of aromatic amino acid biosynthesis in gamma-proteobacteria. *J. Mol. Microbiol. Biotechnol*. **3**, 529–543.
9. Grundy F.J., Henkin T.M. 2003. The T box and S box transcription termination control systems. *Front. Biosci*. **8**, 20–31.