

Research article

Open Access

## Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria

Alexander V Seliverstov<sup>1</sup>, Harald Putzer<sup>2</sup>, Mikhail S Gelfand\*<sup>1</sup> and Vassily A Lyubetsky<sup>1</sup>

Address: <sup>1</sup>Institute for Information Transmission Problems, RAS, Bolshoi Karetny pereulok 19, Moscow, 127994, Russia and <sup>2</sup>Institut de Biologie Physico-Chimique, CNRS UPR9073, 13, rue P. et M. Curie, 75005 Paris, France

Email: Alexander V Seliverstov - slvstv@iitp.ru; Harald Putzer - Harald.Putzer@ibpc.fr; Mikhail S Gelfand\* - gelfand@iitp.ru; Vassily A Lyubetsky - lyubetsk@iitp.ru

\* Corresponding author

Published: 03 October 2005

Received: 01 April 2005

BMC Microbiology 2005, 5:54 doi:10.1186/1471-2180-5-54

Accepted: 03 October 2005

This article is available from: <http://www.biomedcentral.com/1471-2180/5/54>

© 2005 Seliverstov et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Formation of alternative structures in mRNA in response to external stimuli, either direct or mediated by proteins or other RNAs, is a major mechanism of regulation of gene expression in bacteria. This mechanism has been studied in detail using experimental and computational approaches in proteobacteria and Firmicutes, but not in other groups of bacteria.

**Results:** Comparative analysis of amino acid biosynthesis operons in Actinobacteria resulted in identification of conserved regions upstream of several operons. Classical attenuators were predicted upstream of *trp* operons in *Corynebacterium* spp. and *Streptomyces* spp., and *trpS* and *leuS* genes in some *Streptomyces* spp. Candidate leader peptides with terminators were observed upstream of *ilvB* genes in *Corynebacterium* spp., *Mycobacterium* spp. and *Streptomyces* spp. Candidate leader peptides without obvious terminators were found upstream of *cys* operons in *Mycobacterium* spp. and several other species. A conserved pseudoknot (named LEU element) was identified upstream of *leuA* operons in most Actinobacteria. Finally, T-boxes likely involved in the regulation of translation initiation were observed upstream of *ileS* genes from several Actinobacteria.

**Conclusion:** The metabolism of tryptophan, cysteine and leucine in Actinobacteria seems to be regulated on the RNA level. In some cases the mechanism is classical attenuation, but in many cases some components of attenuators are missing. The most interesting case seems to be the *leuA* operon preceded by the LEU element that may fold into a conserved pseudoknot or an alternative structure. A LEU element has been observed in a transposase gene from *Bifidobacterium longum*, but it is not conserved in genes encoding closely related transposases despite a very high level of protein similarity. One possibility is that the regulatory region of the *leuA* has been co-opted from some element involved in transposition. Analysis of phylogenetic patterns allowed for identification of MLI 624 of *M. leprae* and its orthologs as the candidate regulatory proteins that may bind to the LEU element. T-boxes upstream of the *ileS* genes are unusual, as their regulatory mechanism seems to be inhibition of translation initiation via a hairpin sequestering the Shine-Dalgarno box.

## Background

Formation of alternative structures in 5'-leader regions of mRNAs is emerging as a major mechanism of gene regulation. There exist several possible variants of this mechanism whose common feature is the competition between two structures, one of which represses gene expression via premature termination of transcription or inhibition of translation initiation (reviewed in [1-6]). The energetically or kinetically more favourable structure forms by default, whereas the other one is stabilized by binding of a regulatory protein, tRNA, or a small cofactor, or is formed co-transcriptionally, as in classical attenuators.

RNA regulatory elements have been studied mainly in gamma-proteobacteria (*Escherichia coli*) and firmicutes (*Bacillus subtilis*). Computational analysis also has been mainly restricted to proteobacteria [7,8] and firmicutes [9-12]. Recently a new class of regulatory elements, riboswitches, has been described. These elements are highly conserved and were found in all major taxa of bacteria, as well as in some eukaryotes and archaea [13,14]. Comparative genomic analysis has played a major role in the discovery and analysis of T-boxes [9,15] and most riboswitches (reviewed in [4,5]). Several groups performed large-scale search for new RNA regulatory structures [16,17]. Analysis of RNA-based regulation often leads to non-trivial functional assignments for hypothetical genes and filling gaps in metabolic reconstruction (e.g. [11,14,18,19]).

Here we performed comparative analysis of candidate RNA regulatory elements in genomes of Actinobacteria. There are few known attenuators in these genomes. Those that have been experimentally studied are attenuators of the *trp* operons in *Corynebacterium glutamicum* [20] and *Streptomyces venezuelae* [21]. Studies of attenuator-like structures upstream of the *ilvB* and *leuA* genes of *Streptomyces coelicolor* produced somewhat ambivalent results. Indeed, although candidate leader peptides and alternative RNA structures were found upstream of the *ilvB* and *leuA* genes, reminiscent of the classical attenuators, the mutation analysis demonstrated that the regulatory mechanism is not attenuation in the strict sense: mutations in candidate regulatory codons in the leader peptide of the *ilvB* gene had no effect on regulation, and, although mutations in the leader peptide of *leuA* had some effect, it was not consistent with classical attenuation [22]. Computational analysis identified several types of riboswitches: THI-elements [14], RFN-elements [18], B12-elements [19], all of them regulating genes of cofactor metabolism by sequestering the Shine-Dalgarno box and start codon, and interfering with initiation of translation.

## Results and discussion

Following an approach described previously [8], we systematically analysed the upstream regions of amino acid biosynthesis and aminoacyl-tRNA synthetase operons. Candidate regulatory structures were found upstream of genes involved in tryptophan, cysteine, and leucine metabolism. Candidate T-boxes were observed upstream of isoleucyl-tRNA synthetase genes. No conserved structures were observed upstream of genes from other amino acid biosynthesis pathways.

### Tryptophan

The *trp* operons are preceded by classical candidate attenuators in all considered genomes of *Corynebacterium* spp. and *Streptomyces* spp. (Fig. 1). The leader peptides have double or triple repeats of regulatory UGG codons. All terminators are GC-rich and followed by poly-U-tracts. The antiterminator and terminator hairpins in all genomes contain complementary triples gGCC-rGCy-GGCC where absolutely conserved positions are set in capitals. This is analogous to the situation in proteobacteria, where the patterns involved in multiple interactions within attenuators are conserved at large evolutionary distances [8]. In *C. diphtheriae*, candidate attenuators were found upstream of both biosynthetic operons *trpB*<sub>1</sub>*EDGC* and *trpB*<sub>2</sub>*A*. A candidate attenuator was found upstream of the tryptophanyl-tRNA synthetase gene *trpS*<sub>2</sub> in *S. avermitilis*.

### Cysteine

The upstream regions of the *cys* operon in *Mycobacterium* spp. and *Propionibacterium acnes* and the *cbs* gene of *Bifidobacterium longum* contain short open reading frames encoding candidate leader peptides with runs of cysteine codons near the stop codon (Fig. 2a). The upstream regions of *Mycobacterium* spp. are very similar and can be aligned (Fig. 2b). However, they do not contain any conserved hairpins that could serve as terminators of transcription. One possibility is that this region contains rho-dependent terminators similar to the situation in the tryptophanase operon *tna* of *E. coli* [23]. Indeed, *Mycobacterium* spp. have few rho-independent terminators [24,25]. On the other hand, all *Mycobacterium* genomes contain the components of the rho-dependent termination mechanism, *rho*, *nusG*, *nusA*, *nusB*. The region between the candidate leader peptide ORFs and the first genes in the *cys* operons contain polyY motifs that could serve as Rho-binding sites [26-28]. However, these motifs are not conserved, and thus this prediction is rather weak.

The cysteine operons in *M. avium* and *M. leprae* contain additional hypothetical genes, *MAP2122* and *ML0840* respectively, that are 62% identical but have no other reliable homologs.

**a)**

Bacterium	Locus	Gene	Gene coordinates	Protein
<i>C. diphtheriae</i>	NC_002935	<i>trpB1</i>	2456701..2458032	NP_940652
		<i>trpB2</i>	2465139..2466365	NP_940660
<i>C. efficiens</i>	NC_004369	<i>trpE</i>	3052837..3054504	NP_739478
<i>C. glutamicum</i>	NC_003450	<i>trpE</i>	3233404..3234960	NP_602223
<i>S. avermitilis</i>	NC_003155	<i>trpS2</i>	complement(5757496..5758491)	NP_825902
		<i>trpE1</i>	complement(7320283..7322268)	NP_827260
<i>S. coelicolor</i>	NC_003888	<i>trpE</i>	2276703..2278607	NP_626374

**b)**

Bacterium	Operon	Leader peptide
<i>C. diphtheriae</i>	<i>trpB1EGDC1</i>	2456514 -----MNAHN <b>WWW</b> RA----- 2456543
<i>C. diphtheriae</i>	<i>trpB2A</i>	2464983 -----MNAAFK <b>FWW</b> RA----- 2465015
<i>C. efficiens</i>	<i>trpEGDCBA</i>	3052621 VNNFCQSQGTQ <b>WWW</b> RAR----- 3052671
<i>C. glutamicum</i>	<i>trpEGDCBA</i>	3233152 VNNSCLSQSTQ <b>WWW</b> RAN----- 3233199
<i>S. avermitilis</i>	<i>trpS2</i>	5758647 ---MTTRTCTQ <b>QW</b> WAA----- 5758609
<i>S. avermitilis</i>	<i>trpE1</i>	7322414 ---MFAHSIQ <b>NWWW</b> TAHPAAH 7322361
<i>S. coelicolor</i>	<i>trpE</i>	2276540 ---MFAHSTR <b>NWWW</b> TAHPAAH 2276593

**c)**

Bacterium	Operon	Attenuator
<i>C. diphtheriae</i>	<i>trpB1EGDC1</i>	<b>ugguggugg</b> cgcgcu <b>aaacc</b> . <u>g</u> cgggcc. <u>g</u> uuuu...cacgcauuc <u>uuuc</u> .
<i>C. diphtheriae</i>	<i>trpB2A</i>	uu <b>cguggugg</b> cgcgcc <b>uag</b> cagggggccccc <u>uuu</u> ugugugagcauucaccaca
<i>C. efficiens</i>	<i>trpEGDCBA</i>	<b>ugguggugg</b> cgcgcuagau <b>aa</b> gcgggcc <b>cc</b> acggaucaccaaguuguuuu <b>uac</b>
<i>C. glutamicum</i>	<i>trpEGDCBA</i>	<b>ugguggugg</b> cgcgcu <b>aa</b> cu <b>aa</b> gcgagccugacaccu <b>ca</b> aguuguuuu <b>uacuu</b>
<i>S. avermitilis</i>	<i>trpS2</i>	cag <b>uggugggg</b> cgcc <b>uga</b> . <u>c</u> ggcg. <u>g</u> ccguacacacg <u>ua</u> uguacuc.....
<i>S. avermitilis</i>	<i>trpE1</i>	<b>ugguggugg</b> accgcuc <u>accggcg</u> . <u>g</u> cccac <b>uga</b> cugcgcg.....
<i>S. coelicolor</i>	<i>trpE</i>	<b>ugguggugg</b> accgcuc <u>accggcg</u> . <u>g</u> cccac <b>uga</b> cugcgcg.....
<i>S. venezuelae</i>	<i>trpE</i>	<b>ugguggugg</b> accgcuc <u>accggcg</u> . <u>g</u> cccac <b>uga</b> ucgcgcg.....

*C. diphtheriae* .....aac..aggcucgccuugucca....ac.aagcagcgggccuuuuuuuuagc  
*C. diphtheriae* .caacuuuuggaaacacaagcccgcguau.....c.gcgggcuuuuuuuguaau  
*C. efficiens* .acugaagauuu...caaggcucguguaucuuuugacgaagcagcgggccuuuu.gugguua  
*C. glutamicum* ...ugaugaauuuuuuaggcucgu..acuucgucgacgaagaagcgggccuuuu.gugguuuu  
*S. avermitilis* .....aacggcgccgccu.....cgggcgccguucuguuuc  
*S. avermitilis* .acgcaagacuucggaaggccgccc.....gagggcgggccuuucguguuucg  
*S. coelicolor* ..acucaagacuucggaaggccgccc.....gagggcgggccuuucguguuuucg  
*S. venezuelae* acacggaucacacgcacaggccgccc.....gagggcgggccuuucg

**Figure 1**

**Leader peptides and candidate attenuators upstream the *trp* operons in *Corynebacterium* and *Streptomyces* spp.** a) Coordinates and protein identifiers of the first genes in the operons. b) Alignment of the leader peptides. The numbers denote genome positions of the aligned fragments. c) Alignment of the attenuators. Tryptophan and stop codons are shown in bold. The terminator hairpins are highlighted in grey, the antiterminator hairpins are underlined. The alignment contains fragments between the tryptohan codons and the terminator hairpin followed by poly-U-tracts. The numbers denote genome positions of the aligned fragments.

**a)**

Bacterium	Locus	Gene	Gene coordinates	Protein
<i>M. avium</i>	NC_002944	MAP2122	2351330..2352622	NP_961056
<i>M. bovis</i>	NC_002945	<i>cysK1</i>	2586392..2587324	NP_856011
<i>M. tub</i>	NC_002755	<i>cysK</i>	2604640..2605572	NP_336875
CDC1551				
<i>M. tub H37Rv</i>	NC_000962	<i>cysK</i>	2608794..2609726	NP_216850
<i>M. leprae</i>	NC_002677	ML0840	complement(997285..998589)	NP_301634
<i>M. marinum</i>	gnl Sanger_216594 mar22d05.p1c	<i>cysK</i>	complement(136548..137477)	(unfinished)
<i>P. acnes</i>	NC_006085	<i>cysK</i>	1047389..1048324	YP_055674
<i>B. longum</i>	NC_004307	<i>cbs</i>	1006495..1007721	NP_696325

**b)**

Bacterium	Operon	Leader peptide
<i>M. avium</i>	<i>XcysKE</i>	2351124 MQHRLQPRFAPSRLVACCCCCCR 2351177
<i>M. bovis</i>	<i>cysK1E</i>	2586122 MQQAIQLRFILPRRLAVGCCCC--- 2586187
<i>M. tub CDC1551</i>	<i>cysKE</i>	2604371 MQQAIQLRFILPRRLAVGCCCC--- 2604436
<i>M. tub H37Rv</i>	<i>cysKE</i>	2608526 MQQAIQLRFILPRRLAVGCCCC--- 2608591
<i>M. leprae</i>	<i>XcysKE</i>	0998791 MHQSTQPRFVFTRRFTVDCYCRCC- 0998742
<i>M. marinum</i>	<i>cysKE</i>	0138059 MQQAAQLSFVLTRCPAVDCCCC--- 0137994
<i>P. acnes</i>	<i>cysK</i>	1047061 MTSAMMVICRCCC- 1047102
<i>B. longum</i>	<i>cbs</i>	1007876 MQIISCCCR- 1007850

**c)**

	RBS	Start
<i>M. avium</i>	uauaguggugac	<b>aug</b> caacaccgccuacagccgcgcuuu
<i>M. bovis, tub</i>	uauaguggggccc	<b>aug</b> caacaggccauacagcugcgcguuu
<i>M. leprae</i>	uauaguggaccu	<b>aug</b> caucaguccacacagccacgcguuu
<i>M. marinum</i>	uauaguagagcc	<b>aug</b> caacaggccgcacagcugagcguuu

	Cys tract
<i>M. avium</i>	gccccgucgcgcgucgcccugucguggcc <b>uguuguugcuguuguugc</b> gcu
<i>M. bovis, tub</i>	auccucccgcgccgcccucgcccgugggg <b>uguuguuguugu</b> .....
<i>M. leprae</i>	gucuuuacgcgcgcccguuuaccguggac <b>uguuauugucgcuguugc</b> ...
<i>M. marinum</i>	guccucacgcgcgucgccccgcccguggac <b>uguuguuguugcugu</b> .....

	Stop and putative Rho binding site
<i>M. avium</i>	<b>ug</b> AUUUCCgcaaGCCCUCugacgcuguagaaAUCCCCgcgucGCCCCUgcccc
<i>M. bovis, tub</i>	<b>ug</b> AUCCUg.gcguccacagcaAUCCUcgcGCUCUgcccc
<i>M. leprae</i>	<b>ug</b> AUCCUgac.ACCUUUaacGCUCUCagcaaaucauucacGUUCUCgccua
<i>M. marinum</i>	<b>ug</b> AUCCUgac.gcguucugaccguccaguaaucgucGCCUCUgucgccucaugg

**Figure 2**  
**Leader peptides upstream the cys operons in Mycobacterium spp. and P. acnes and cbs operon in B. longum.** a) Coordinates and protein identifiers of the first genes in the operons. b) Alignment of the leader peptides. The numbers denote genome positions of the aligned fragments. c) DNA alignment of the leader peptide genes. Start, cysteine and stop codons are shown in bold; candidate Rho-binding sites are shown in capitals.

**Leucine**  
 The upstream regions of the *ilvB* genes (operons *ilvBNC*, *ilvBHC*, *ilvBserA<sub>1</sub>*) in *Corynebacterium*, *Mycobacterium*, *Streptomyces* species contain short ORFs with runs of iso-leucine, valine and leucine codons overlapping the candi-

date terminator hairpins followed by polyU-runs (Fig. 3). However, the exact mode of regulation is not clear, as experimental substitution of possible regulatory codons upstream of the *ilvBNC* operon in *S. coelicolor* had no effect on regulation or expression of *ilvB* [23].

**a)**

Bacterium	Locus	Gene	Gene coordinate	Protein
<i>C. diphtheriae</i>	NC_002935	<i>ilvB</i>	1082013..1083971	NP_939459
<i>C. efficiens</i>	NC_004369	<i>ilvB</i>	1432330..1434327	NP_737975
<i>C. glutamicum</i>	NC_003450	<i>ilvB</i>	1338131..1340011	NP_600493
<i>M. tuberculosis</i> H37Rv	NC_000962	<i>ilvB</i>	complement(3361127..3362983)	NP_217519
<i>M. tuberculosis</i> CDC1551	NC_002755	<i>ilvB</i>	complement(3355506..3357362)	NP_337598
<i>M. bovis</i>	NC_002945	<i>ilvB1</i>	complement(3317745..3319601)	NP_856673
<i>M. leprae</i>	NC_002677	<i>ilvB</i>	complement(2044335..2046212)	NP_302166
<i>M. avium</i>	NC_002944	<i>ilvB1</i>	complement(3379032..3380900)	NP_961972
<i>M. marinum</i> gnl Sanger_216594 mar755h11.p2k1114			complement(164709..166565)	(unfinished)
<i>S. avermitilis</i>	NC_003155	<i>ilvB</i>	complement(3354433..3356283)	NP_823909
<i>S. coelicolor</i>	NC_003888	<i>ilvB</i>	6003117..6004958	NP_629647

**b)**

Bacterium	Operon		Leader Peptide	
<i>C. diphtheriae</i>	<i>ilvBHC</i>	1081747	MNIIRLVVITTRRLP	1081791
<i>C. efficiens</i>	<i>ilvBHC</i>	1432212	MTSIRPVVIVAARRLP-	1432259
<i>C. glutamicum</i>	<i>ilvBHC</i>	1337840	MTIIRLVVVTARRLP	1337884
<i>M. tuberculosis</i> H37Rv	<i>ilvBNC</i>	3363152	MDKAGKPGMLVVIGRRVGA	3363096
<i>M. tuberculosis</i> CDC1551	<i>ilvBNC</i>	3357528	MDKAGKPGMLVVIGRRVGA	3357472
<i>M. bovis</i>	<i>ilvB1NC</i>	3319767	MDKAGKPGMLVVIGRRVGA	3319711
<i>M. leprae</i>	<i>ilvBNC</i>	2046378	MLVVICQRVGG	2046346
<i>M. avium</i>	<i>ilvB1N</i>	3381051	MLVVI-RRVGA	3381022
<i>M. marinum</i>	<i>ilvB</i>	166742	MDTAGTPGKLVVLGRRVVA	166686
<i>S. avermitilis</i>	<i>ilvBNC</i>	3356481	MRTRILVLGKRVG	3356443
<i>S. coelicolor</i>	<i>ilvBNC</i>	6002909	MRTRILVLGKRVG	6002947

**c)**

Bacterium	Terminator
<i>C. diphtheriae</i>	aaaagcg . . . cccucgacag . . . caccacacaugcugagcgggggcuuuccuuau
<i>C. efficiens</i>	caa . gcg . . . cccucgacaguaccaccacagugcuguuucgaggggcuuuguugu .
<i>C. glutamicum</i>	caa . gcg . . . cccucgacaacacucaccacaguguuggaacgaggggcuuucuuuguu
<i>M. tuberculosis</i>	caacgcg . . . acccucgugcagcagc . . . . . ugagcuggcga . ggguuuuuuuuuu
<i>M. bovis</i>	caacgcg . . . acccucgugcagcagc . . . . . ugagcuggcga . ggguuuuuuuuuu
<i>M. leprae</i>	caacgcgcaaccucgugcagcuag . . . . . ucagcugucga . ggguuuuuuuguu
<i>M. avium</i>	caacgcgcaaccucgugcagcaca . . . . . agcugucg . ggguuuuuuuguu
<i>M. marinum</i>	caacgcgcaaccucgugcagcag . . . . . cugagcugacg . ggguuuuuuuguu
<i>S. avermitilis</i>	cggcgcgcuccccucgcuugcc . . . . . ucacggcacgaggggguuuuuuguu
<i>S. coelicolor</i>	cgacgcgcuccccucgcuugcc . . . . . uuacggcacgaggggguuuuuuguu

**Figure 3**  
**Candidate leader peptides and terminators upstream the *ilv* operon in Actinobacteria.** a) Coordinates and protein identifiers of the first genes in the operons. b) Alignment of the leader peptides. The numbers denote genome positions of the aligned fragments. c) Alignment of the terminators. The terminator hairpins are highlighted in grey.

Classical candidate attenuators were found upstream of *leuS* (leucyl-tRNA-synthetase) in *S. avermitilis* and *S. coelicolor*. Each of them contains an ORFs encoding the leader peptide, as well as the antiterminator and terminator hairpins (Fig. 4).

Sequences upstream of the isopropylmalate synthase genes *leuA* contain a number of candidate regulatory sequences, together named the LEU element (Fig. 5, 6). Firstly, there is an upstream ORF encoding a candidate leader peptide with a run of leucine codons (Fig. 7).

**a)**

Bacterium	Locus	Gene	Gene coordinates	Protein
<i>S. avermitilis</i>	NC_003155	<i>leuS</i>	6661895..6664783	NP_826665
<i>S. coelicolor</i>	NC_003888	<i>leuS</i>	complement(2775536..2778436)	NP_626809

**b)**

		<b>M R A V R L L L S E P R</b>	
<i>S. avermitilis</i>	6661741	<b>aug</b> cgugccguacgccc <b>uucugc</b> uuagcgagccgcgc <b>ug</b> aucagcccagaccacugacga	
<i>S. coelicolor</i>	2778624	<b>aug</b> cgugccguacgccc <b>uucugc</b> uuagcgagccgcgc <b>ug</b> aucaguccccgacccccggucgu	
<i>S. avermitilis</i>	. .uuc	<u>guggucggaucggcgcgggcg</u> ccccu <b>ccugucgagggg</b> uuuuuucauu	6661852
<i>S. coelicolor</i>	aguccgguggcggaucggcgcgggcgccccu <b>ccugucgagggg</b> uuuuuucauu		2778510

**Figure 4**  
**Candidate attenuators upstream the *leuS* operon in *Streptomyces* spp.** a) Coordinates and protein identifiers of the *leuS* genes. b) Alignment of the attenuators. Start, leucine and stop codons are shown in bold. The terminator hairpins are highlighted in grey, the antiterminator hairpins are underlined. The alignment contains fragments between the leader peptide ORFs and the terminator hairpin followed by poly-U-tracts.

Secondly, this region may fold into a pseudoknot with an additional stem at its base formed by pairing of the leucine codon run with the Shine-Dalgarno box of the *leuA* gene (Fig. 5, 8). Finally, the same region may form an alternative hairpin with the same base stem (Fig. 6).

A similar pseudoknot was found in *B. longum* within a gene encoding a transposase. The latter is homologous to the IS1554 transposase of *M. tuberculosis* and *M. bovis* (66% identity), a putative transposase in *C. efficiens* (40% identity), putative IS256 family transposases of *S. avermitilis* (31% identity), hypothetical protein MAP2274 of *M. avium* (29% identity), and some other putative transposases from *B. longum*, *C. efficiens*, *M. tuberculosis*, *M. bovis*, *R. xylanophilus*, *S. avermitilis*, *S. coelicolor* (Fig. 9a). However, only the *B. longum* transposase contains a fragment that may fold into the pseudoknot (Fig. 9b), whereas other transposases, although highly similar on the protein level in the corresponding region, contain a number of non-complementary mismatches in synonymous codon positions and thus have lost the pseudoknot folding potential.

**T-boxes**

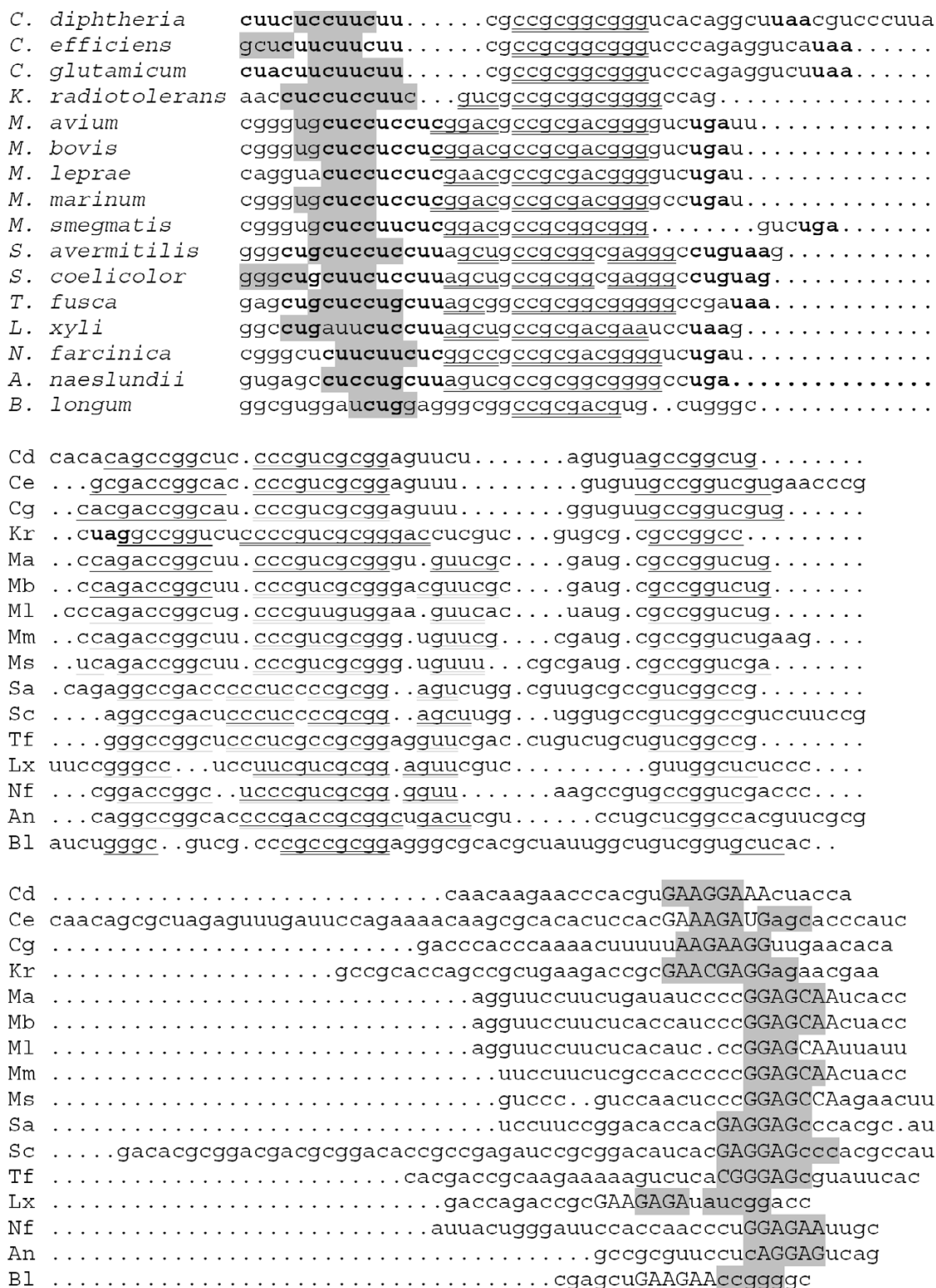
Candidate T-box structures were found upstream of the *ileS* genes from several Actinobacteria. They are unusual, as instead of terminators, they contain hairpins sequestering the Shine-Dalgarno boxes of the *ileS* genes (Fig. 10).

Thus it is likely that the regulatory mechanism involves inhibition of translation initiation. To our knowledge, this is the first example of a T-box acting on the level of translation.

**Conclusion**

Candidate regulatory elements were found upstream of genes involved in the tryptophan, cysteine and branched chain amino acids metabolism. No conserved RNA regulatory structures were observed upstream of histidine, threonine, phenylalanine, tyrosine, arginine, lysine, methionine operons, although orthologous genes involved in the latter pathways are regulated on the RNA level in other species: methionine and lysine by the S-box and L-box riboswitches respectively [3-5], histidine, threonine and phenylalanine by attenuators [7,8], tyrosine and arginine by T-boxes [12].

Attenuators of the classical type were observed upstream of the aminoacyl-tRNA-synthetase genes *trpS* and *leuS* in some *Streptomyces* genomes, similar to those observed in gamma-proteobacteria, (e.g. the *pheST* operon) [7]. In contrast, in Firmicutes, most aminoacyl-tRNA-synthetase genes are regulated by tRNA-dependent antitermination (T-boxes) and none by classical attenuation [2,9,15]. No classical T-boxes were found in Actinobacteria, but unusual T-boxes, possibly regulating initiation of translation,



**Figure 5**  
**Alignment and RNA secondary structures of the *leuA* upstream regions (LEU elements).** The stem at the base is highlighted in grey, helices forming the pseudoknot are underlined and double underlined, leucine and stop codons are set in bold, the candidate Shine-Dalgarno boxes of the *leuA* are set in capitals. The last sequence is that of the transposase from *B. longum* (see the text). Sequences for *M. bovis* (Mb) and *M. tuberculosis* spp. (Mt and Rv) coincide.

*C. diphtheria* cuucuccuucuu.....cgccgcgggcgggucacagggcuuaacgucccuua  
*C. efficiens* gcucuuucuuuu.....cgccgcgggcggguccagaggucuuuaa.....  
*C. glutamicum* cuacuucuuuu.....cgccgcgggcggguccagaggucuuuaa.....  
*K. radiotolerans* aaccuccuccuuc...gucgcccggcggggcccag.....  
*M. avium* cgggugcuccuccucggacgcccgcgacggggucugauu.....  
*M. bovis* cgggugcuccuccucggacgcccgcgacggggucugau.....  
*M. leprae* cagguacuccuccucgaacgcccgcgacggggucugau.....  
*M. marinum* cgggugcuccuccucggacgcccgcgacggggcccugau.....  
*M. smegmatis* cgggugcuccuucucggacgcccgcggcggg.....gucuga.....  
*S. avermitilis* gggcugcuccuccuuagcugcccgcggcgagggccuguaag.....  
*S. coelicolor* gggcugcuccuccuuagcugcccgcggcgagggccuguaag.....  
*T. fusca* gagcugcuccucgcuuagcggcccgcggcgggggccgauaa.....  
*L. xyli* ggccugauucuccuuagcugcccgcgacgaauccuaag.....  
*N. farcinica* cgggcucuucuucggcccgcgcccgcgacggggucugau.....  
*A. naeslundii* gugagccuccucgcuuagcugcccgcggcggggcccuga.....  
*B. longum* ggcguggaucuggagggcgggcccgcgacgug..cugggc.....

Cd cacacagccggcuc.cccgucgcggaguucuu.....aguguagccggcug.....  
Ce ...gcgaccggcac.cccgucgcggaguuu.....guguugccggucgugaaccgg  
Cg ..cacgaccggcau.cccgucgcggaguuu.....gguguugccggucgug.....  
Kr ..cuaggccggucucccccgucgcgggaccucguc...gugcg.cgccggcc.....  
Ma ..ccagaccggcuu.cccgucgcggggu.guucgc...gaug.cgccggucug.....  
Mb ..ccagaccggcuu.cccgucgcggggacguucgc...gaug.cgccggucug.....  
Ml .cccagaccggcug.cccguuguggaa.guucacu...aug.cgccggucug.....  
Mm ..ccagaccggcuu.cccgucgcgggg.uguucg...cgaug.cgccggucugaag...  
Ms ..ucagaccggcuu.cccgucgcgggg.uguuu...cgcgaug.cgccggucga.....  
Sa .cagagccgacccccuccccgccggag..ucugg.cguugcgcgucggccc.....  
Sc ....aggccgacuccuccccgccggag..cuugg...uggugcgcgucggcccucuccg  
Tf ....gggcccggucuccccgccggagguucgac.cugucugcugucggccc.....  
Lx uuccgggccc...uccuucgucgcggag.uucguc.....guuggcucucc...  
Nf ...cggaccggc...uccgucgcgggg.uu.....aagccgugccggucgacc...  
An ...cagccggcaccccgacccggucgacucgu.....ccugcucggccacguucgag  
Bl aucugggc..gucg.cccgccccggagggcgacgcuauggcugucgggucac..

Cd .....caacaagaaccacguGAAGGAAAcuacca  
Ce caacagcgcuagaguuugauucagaaaacaagcgcacacuccacGAAAGAUgagcaccac  
Cg .....gaccaccccaaaacuuuuAAGAAGGuugaacaca  
Kr .....gccgcaccagccgcgugaagaccgcGAACGAGGagaacgaa  
Ma .....agguuccuucugauaucccGGAGCAAucacc  
Mb .....agguuccuucaccaucccGGAGCAAucacc  
Ml .....agguuccuucacacac.cccGGAGCAAuuau  
Mm .....uuccuucucgccacccccGGAGCAAucacc  
Ms .....guccc..guccaacuccGGAGCAAagaacuu  
Sa .....uccuuccgacaccacGAGGAGcccacgc.au  
Sc ....gacacgaggacgacgaggacaccgcccagauccgaggacauacGAGGAGcccacgcca  
Tf .....cacgaccgcaagaaaagucucaCGGGAGcguauucac  
Lx .....gaccgacccgcGAAGAGAuauccgacc  
Nf .....auuacugggauuccaccaaccuGGAGAAuugc  
An .....gccgcuuccuAGGAGucag  
Bl .....cgagcuGAAGAAccggggc

**Figure 6**  
**Alternative RNA secondary structure in LEU elements.** The stem at the base is highlighted in grey, two internal helices are underlined and double underlined, other notation as in Fig 5.



a)

Bacterium	Locus	Gene	Gene coordinates	Protein
<i>C. diphtheria</i>	NC_002935	<i>leuA</i>	complement(228555..230372)	NP_938656
<i>C. efficiens</i>	NC_004369	<i>leuA</i>	complement(233589..235439) (adding 105 nucleotides)	NP_736826
<i>C. glutamicum</i>	NC_003450	<i>leuA</i>	complement(266151..268001)	NP_599502
<i>K. radiotolerans</i>	AAEF02000060	<i>leuA</i>	complement(3238..4965)	EAM73829
<i>M. avium</i>	NC_002944	<i>leuA</i>	333789..335633	NP_959246
<i>M. bovis</i>	NC_002945	<i>leuA</i>	4091088..4093193	NP_857375
<i>M. tub CDC1551</i>	NC_002755	<i>leuA</i>	4145949..4147928	NP_338367
<i>M. tub H37Rv</i>	NC_000962	<i>leuA</i>	4153737..4155671	NP_218227
<i>M. leprae</i>	NC_002677	<i>leuA</i>	2754640..2756463	NP_302512
<i>M. marinum</i>	gnl Sanger_216594 mar428a07.p1k		192528..194345	(unfinished)
<i>M. smegmatis</i>	gnl TIGR_246196 contig:3563:m_smeigmatis		6334690..6336495	(unfinished)
<i>S. avermitilis</i>	NC_003155	<i>leuA2</i>	6774328..6776049	NP_826778
<i>S. coelicolor</i>	NC_003888	<i>leuA</i>	complement(2725480..2727201)	NP_733575
<i>T. fusca</i>	NZ_AAAQ02000002	<i>leuA</i>	349237..350943 (adding 27 nucleotides)	ZP_00293601
<i>L. xyli</i>	NC_006087	<i>leuA</i>	complement(1501628..1503400)	YP_062368
<i>N. farcinica</i>	NC_006361	<i>leuA</i>	complement(322994..324787)	YP_116514
<i>A. naeslundii</i>	gnl TIGR_240017 contig:1063:a_naeslundii		594374..596211	(unfinished)

b)

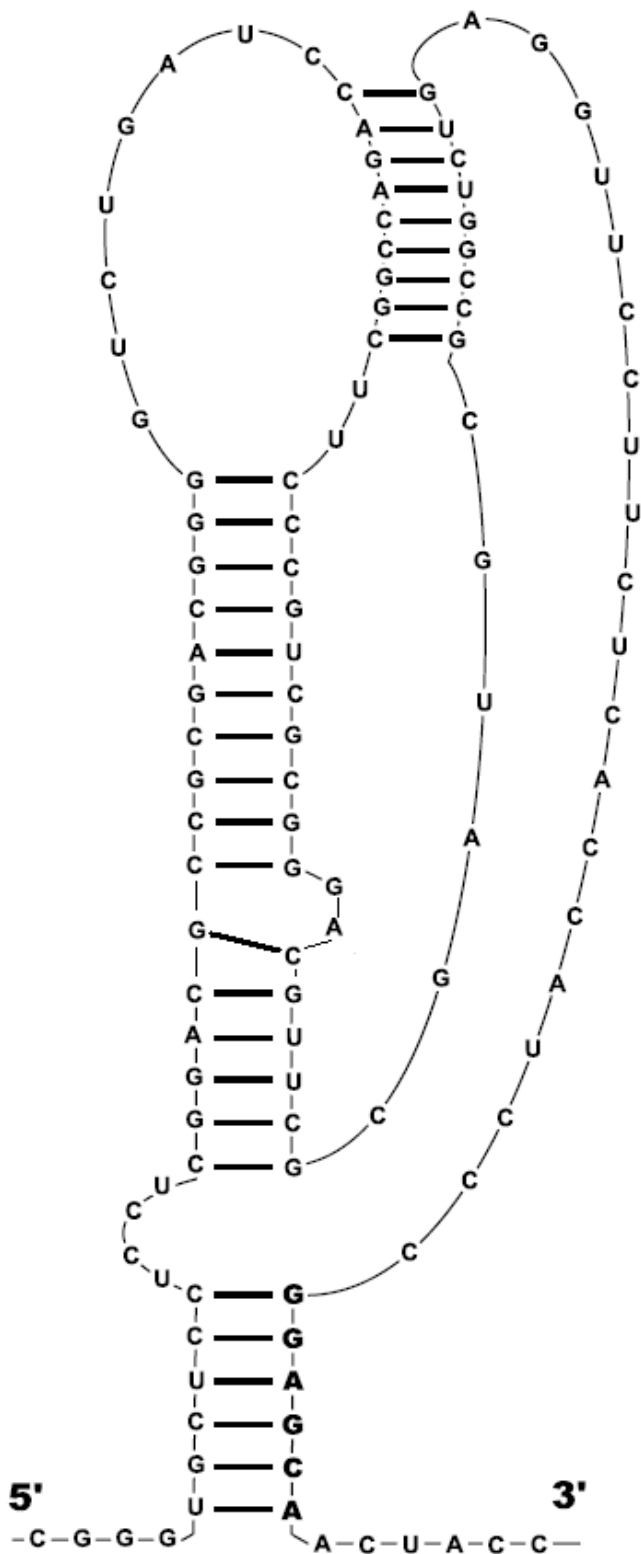
Bacterium	Leader peptide
<i>C. diphtheria</i>	230506 MNRANLLLLRRGGSQA- 230459
<i>C. efficiens</i>	235612 MFSSHERSALLLRGGSQRS 235553
<i>C. glutamicum</i>	268124 MTSRANLLLLRRGGSQRS 268095
<i>K. radiotolerans</i>	5097 VARLENLLLLRRRGAS- 5050
<i>M. avium</i>	333705 VADVQRVLLLGRRDGV-- 333752
<i>M. bovis</i>	4090959 VLHVQRVLLLGRRDGV--4091006
<i>M. tub CDC1551</i>	4145866 VLHVQRVLLLGRRDGV--4145913
<i>M. tub H37Rv</i>	4153611 VLHVQRVLLLGRRDGV--4153658
<i>M. leprae</i>	2754521 VQVLLLERRDGV--2754559
<i>M. marinum</i>	192399 VLCVQRVLLLGRRDG--- 192443
<i>M. smegmatis</i>	6334564 VLGVQRVLLLGRRGV--6334611
<i>S. avermitilis</i>	6774199 MRFGLLLS CRGEGEGL-6774243
<i>S. coelicolor</i>	2727361 MRFGLLLS CRGEGEGL-2727317
<i>T. fusca</i>	349104 MLRELLLSGRGGGR- 349148
<i>L. xyli</i>	1503533 MRVTLGLVYGLILLS CRDES--1503474
<i>N. farcinica</i>	324906 MQRALLLGRRDGV-- 324868
<i>A. naeslundii</i>	594266 VSLLLSRRGGA-- 594298

**Figure 7**  
Candidate leader peptides in the LEU elements.

were observed upstream of the *ileS* genes in several genomes.

Despite the presense of conserved leader peptides upstream of some cysteine and leucine operons, the mode

of regulation is unknown, as other attenuator elements are missing. One possible explanation is that attenuation of the *cys* operons in *Mycobacterium* spp. and *P. acnes* and the *cbs* operon in *B. longum* involves Rho-dependent termination, similar to the *tna* operon of *E. coli* [23,29].



**Figure 8**  
**Candidate RNA pseudoknot upstream of the *leuA* operon in *M. bovis*.** The corresponding alignment is given Fig. 5. Boldface: the candidate Shine-Dalgarno box.

The most interesting case seems to be that of the *leuA* genes. The upstream regions of these genes contain several conserved elements (referred to as the LEU element) that can be interpreted in different ways. There are some architectural similarities with riboswitches, in particular, a compact structure with a stem at the base [5,30,31]. The latter is formed by interaction of a run of leucine codons and the Shine-Dalgarno box. Indeed, Actinobacteria seem to be the only taxonomic group where the base stems of riboswitches directly overlap the translation initiation site, without additional regulatory hairpins [5]. However, the LEU element differs from all known riboswitches, as the alignment of LEU elements does not contain conserved unpaired nucleotides that would be involved in tertiary interactions and form the ligand-binding pocket, as in the purine riboswitches whose spatial structure has been resolved [30,31] and in other riboswitches [5]. Thus direct binding of a small molecule to LEU elements seems unlikely. On the other hand, there is experimental evidence that mutations in the leucine codons do not influence the regulation [22] and thus classical attenuation involving translation of a leader peptide also is an unlikely mechanism of regulation.

The above considerations make it likely that the LEU element is a binding site of some regulatory protein. To test for this possibility, we compared the pattern of phylogenetic distribution of LEU elements to phylogenetic distributions of all actinobacterial genes. The closest phylogenetic pattern was observed for orthologs of ML1624 from *M. leprae*: homologs of this protein with E-values  $<10^{-170}$  were found in all genomes containing LEU elements, but not outside Actinobacteria. The only unexplained fact is the presence of a homolog with the E-value  $\sim 10^{-108}$  in *P. acnes*, which does not have a LEU element. The structure of the ML1624 protein is consistent with an RNA-binding regulatory role, as the protein contains an N-terminal DEAD-box helicase domain (ProFam family PF00270, E-value  $3.6 \cdot 10^{-6}$ ) that may be involved in unwinding of nucleic acids.

An additional enigma is the presence of a LEU element-like sequence within a transposase gene. On the other hand, it may be a clue to the origin of LEU elements. One possibility is that the *B. longum* transposase represents an ancestral state where the LEU element was involved in maintenance or regulation of transposition. Situations when a regulatory site occurs within a regulatory and/or regulated gene are not very common, but they happen in mobile elements [32]. Other transposase genes may have lost the ability to form this structure due to mutations; notably, the protein sequence has not changed much (Fig. 9), as most mutations occurred in synonymous codon positions. A plausible scenario is that the transposase gene was inserted upstream of the *leuA* gene in the ancestral

a)

Bacterium	Locus	Coordinates	Protein
<i>B. longum</i>	NC_004307	2124903..2126108	NC_004307
<i>M. bovis</i>	NC_002945	complement(1025963..1027282)	NP_854601
<i>M. tuberculosis</i> CDC1551	NC_002755	complement(1025510..1026829)	NP_335380
<i>C. efficiens</i>	NC_004369	complement(1561522..1562694)	NP_738106

b)

```

B. longum
M. bovis
M. tuberculosis CDC1551
C. efficiens

M D A A Q V I E P A H A G Q D V D E A A V A A
M D A A Q V I E P A H A G Q D V D E A A V A A

B1 MAKEKGLDLTGPDGLLKQFTKSVLETAALDEEMTEHLGR**AKHKKSKDGRAANTRNGTTAKTVVTDVSVGPVGVIEVPRDRDGS
Mb RELSGAERALVGDVLRQARAEGVALTGPDGLLKALTKTVLEAALQEEMTEHLGY***DRHAAAGRGSNGSRNRSRNNKVI TDACGQVEI AVPRDRNGT
Mt RELSGAERALVGDVLRQARAEGVALTGPDGLLKALTKTVLEAALQEEMTEHLGY***DRHAAAGRGSNGSRNRSRNNKVI TDACGQVEI AVPRDRNGT
Ce MNAEMDAHLGYGHSRDRDGKTAAGQGNHRNGYYPK*RVDSNYGPI DVAVPRDRNGS
*****EM**HLG*****N*RN*****K*****G*****VPRDR*GS

B1 FEPVVVKRKRRLPGVDEVVLSLYARGLTTGEISAHFQEIYGADVSRVETVSRITRIVVAEKDEWCSRPLDRVYAAVFIDATVVKVRDG*QVANRAFYVAV
Mb FEPVIVGKRKRRTDVRVVLVSLYAKGLTTGEIAAHFADVYGVSVSKDTISRITDRVIEEMQAWWSRPLEKVYAAVFIDAIMVKIRDG*QVRNRPVYAAI
Mt FEPVIVGKRKRRTDVRVVLVSLYAKGLTTGEIAAHFADVYGVSVSKDTISRITDRVIEEMQAWWSRPLEKVYAAVFIDAIMVKIRDG*QVRNRPVYAAI
Ce FLPTMVPKGRRLTDVDDMI*LYAGGMTVRDIQHMHITSMGVDISHETISAITDAVLDEVMWQNRQLDDFYVPIFLDALRIKVRDGGRRVNNKSVYLLAI
F*P**V*K**RR**VD*****LYA*G*T**I**H*****G**S**TIS*IT*AV**E**W**R*L**Y**F*DA**K*RDG**V*N**Y*A*

=====
B1 GVDLEGGRDVLGIWASPA*AEGARYWLSVLTTELKNRGVDDVFFLIDGLKGLPDAVGAVWPLAIVQTCVVHLLRNTFRYASKKDWDAIKRDVKPIYTAAS
Mb GVDLDGHKIDILGMWAGEGDGESAKFWLAVLTELNRNGVKDIFFLVCDGLKGLPDSVSAAFPLATVQTCI IHLIRNTFRYASRKYWDKISVDLKPITYTAAS
Mt GVDLDGHKIDILGMWAGEGDGESAKFWLAVLTELNRNGVKDIFFLVCDGLKGLPDSVSAAFPLATVQTCI IHLIRNTFRYASRKYWDKISVDLKPITYTAAS
Ce GVDIDGIKHILGIWLAKE**EGASFWANVCANLATRGVQDVFIVCCDGLKGLPQAVEATWPDMSVQTCVVHLIRAAANRWVYAGDRKAVSAQLRKIYTAAPT
GVD**G***L*G**W*****E*A*F**V**L**R*G**V**D**F***CDGLKGLP**V*A**P***VQTC**HLIR***R*****IYTA**

B1 **AAAAAARDAMLDKWEARYPAIRRLWMDAWERFIPFLDYDVEIRRVICTTNAIESLNARFKRSIRARGHFDPDEQAALKCMYLTVRSLDPTGKGRIRWS
Mb **AAEARLRYEEFAEKWGPYPATIRLWDSAWEEFIPFLDYDVEIRRVPCSTNAIESLNARYRRAVRARGHFDPNEQSALKTLVLRSLDPKGTGQTKWA
Mt **AAEARLRYEEFAEKWGPYPATIRLWDSAWEEFIPFLDYDVEIRRVPCSTNAIESLNARYRRAVRARGHFDPNEQSALKTLVLRSLDPKGTGQTKWA
Ce EDTAIAALEEFAESELGVK*YPSAKVWRDADRFPFLQFPMPARKVIYTTNSIESMNNELRKATRNRVQFTNDESAIKTLWLMICNIEDKRAAKRAKQ
***A*A*****YP*****W**AW**FIPFL*****R*V**TN*IES*N*****R***F*****AIK**L*****

B1 ARWKPALNAFAITFADRPSEGTQQ
Mb VRWKPALNALAITFADRMPEAEBER
Mt VRWKPALNALAITFADRMPEAEBER
Ce GKRVAASSGRLIEGRKVANWKAQINQMAVAFPPDRFEAYL
*****I*****
    
```

c)

```

B. longum      G V D L E G G R D V L G I W A S P A * A E
               ggcgtggatctggagggcgccgcgacgtgctgggcatctggcgctcgccgcc**gcgag
M. bovis      G V D L D G H K D I L G M W A G E G D G E
               ggcgtcgacctcgacggccacaaggacatcctggggatgtggcgccggaaggcgacggtgag
M. tuberculosis G V D L D G H K D I L G M W A G E G D G E
               ggcgtcgacctcgacggccacaaggacatcctggggatgtggcgccggaaggcgacggtgag

B. longum      G A R Y W L S V L T E L K N R G
               ggcgcacgctattggctgctcggtgctcaccgagctgaagaaccggggc
M. bovis      S A K F W L A V L T E L R N R G
               tcagccaaatthtggctggcagtgctcaccgaactgcgcaatcgtggg
M. tuberculosis S A K F W L A V L T D L R N R G
               tcagccaaatthtggctggcagtgctcaccgacctgcgcaatcgtggg
    
```

**Figure 9**  
**Multiple alignments of transposases.** a) Coordinates and protein identifiers of putative transposases. b) Protein alignment. The fragment marked by the double line above corresponds to the *B. longum* fragment homologous to candidate pseudoknot and shown in the last line of Fig. 5. c) Nucleotide alignment of the region shown by the double line in (b).



actinobacterial genome. The main fraction of the coding sequence was subsequently deleted, whereas the structural element was co-opted for regulation of the downstream *leuA* gene.

## Methods

Genomes of Actinobacteria *Actinomyces naeslundii* (An), *Bifidobacterium longum* (Bl), *Corynebacterium diphtheriae* (Cd), *Corynebacterium efficiens* (Ce), *Corynebacterium glutamicum* (Cg), *Kineococcus radiotolerans* (Kr), *Leifsonia xyli* (Lx), *Mycobacterium avium* (Ma), *Mycobacterium bovis* (Mb), *Mycobacterium leprae* (Ml), *Mycobacterium marinum* (Mm), *Mycobacterium smegmatis* (Ms), *Mycobacterium tuberculosis* (Rv and Mt), *Nocardia farcinica* (Nf), *Propionibacterium acnes* (Pa), *Rubrobacter xylanophilus* (Rx), *Streptomyces avermitilis* (Sa), *Streptomyces coelicolor* (Sc), *Thermobifida fusca* (Tf), *Tropheryma whipplei* (Tw) were downloaded from the NCBI web site. We also used sequences of *Streptomyces venezuelae* (Sv) from [21].

Candidate operons were defined as chains of genes transcribed in the same direction with intergenic regions not exceeding 150 nucleotides. Multiple alignments of genes were used to verify and, if necessary, revise annotated gene starts [33]. The revisions included adding 105 nucleotides (35 codons) to the *leuA* gene from *C. efficiens*, adding 27 nucleotides (9 codons) of the *leuA* gene from *T. fusca*, and removing 147 nucleotides (49 codons) of the *ileS* gene from *C. efficiens*.

RNA sequence and structure alignments were constructed using MultAlign (A.A. Mironov, personal communication) and the program GL [34]. Search for RNA structural patterns was performed using the PAT program (A.V. Seliverstov, unpublished). Search for conserved sequence fragments was done using the CLIQUE program [35]. Multiple protein sequence alignments were constructed using MultAlign.

## Authors' contributions

AVS and VAL developed algorithms. AVS wrote the programs and performed sequence analysis. HP and AVS identified translational T-boxes. AVS, VAL, and MSG analyzed LEU elements. AVS and MSG performed functional annotation and wrote the paper. VAL and MSG conceived and supervised the project.

## Acknowledgements

We are grateful to Andrei Mironov for the MultAlign program and to K.Yu. Gorbunov for useful discussion. This work was partially supported by grants from ISTC (2766), Howard Hughes Medical Institute (55000309), Russian Academy of Sciences (programs "Molecular and Cellular Biology" and "Origin and Evolution of the Biosphere"), Fund for Support of Russian Science.

## References

- Henkin TM, Yanofsky C: **Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions.** *Bioessays* 2002, **24**:700-707.
- Grundy FJ, Henkin TM: **The T box and S box transcription termination control systems.** *Front Biosci* 2003, **8**:d20-31.
- Grundy FJ, Henkin TM: **Regulation of gene expression by effectors that bind to RNA.** *Curr Opin Microbiol* 2004, **7**:126-131.
- Mandal M, Breaker RR: **Gene regulation by riboswitches.** *Nat Rev Mol Cell Biol* 2004, **5**:451-463.
- Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS: **Riboswitches: the oldest mechanism for the regulation of gene expression?** *Trends in Genetics* 2004, **20**(1):44-50.
- Yanofsky C: **The different roles of tryptophan transfer RNA in regulating trp operon expression in E. coli versus B. subtilis.** *Trends in Genetics* 2004, **20**:367-74.
- Panina EM, Vitreschak AG, Mironov AA, Gelfand MS: **Regulation of aromatic amino acid biosynthesis in gamma-proteobacteria.** *J Mol Microbiol Biotechnol* 2001, **3**:529-543.
- Vitreschak AG, Lyubetskaya EV, Shirshin MA, Gelfand MS, Lyubetsky VA: **Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis.** *FEMS Microbiology Letters* 2004, **234**:357-370.
- Grundy FJ, Henkin TM: **Conservation of a transcription antitermination mechanism in aminoacyl-tRNA synthetase and amino acid biosynthesis genes in gram-positive bacteria.** *J Mol Biol* 1994, **235**:798-804.
- Grundy FJ, Henkin TM: **The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria.** *Mol Microbiol* 1998, **30**:737-749.
- Murphy BA, Grundy FJ, Henkin TM: **Prediction of gene function in methylthioadenosine recycling from regulatory signals.** *J Bacteriol* 2002, **184**:2314-2318.
- Panina EM, Vitreschak AG, Mironov AA, Gelfand MS: **Regulation of biosynthesis and transport of aromatic amino acid in low-GC Gram-positive bacteria.** *FEMS Microbiol Lett* 2003, **222**:211-220.
- Sudarsan N, Barrick JE, Breaker RR: **Metabolite-binding RNA domains are present in the genes of eukaryotes.** *RNA* 2003, **9**:644-7.
- Rodionov DA, Vitreschak AA, Mironov AA, Gelfand MS: **Computational analysis of thiamin regulation in bacteria: Possible mechanisms and new THI-element-regulated genes.** *J Biol Chem* 2003, **277**:48949-48959.
- Henkin TM, Glass BL, Grundy FJ: **Analysis of the Bacillus subtilis tyrS gene: conservation of a regulatory sequence in multiple tRNA synthetase genes.** *J Bacteriol* 1992, **174**:1299-1306.
- Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, Wickiser JK, Breaker RR: **New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control.** *Proc Natl Acad Sci USA* 2004, **101**:6421-6426.
- Abreu-Goodger C, Ontiveros-Palacios N, Ciria R, Merino E: **Conserved regulatory motifs in bacteria: riboswitches and beyond.** *Trends Genet* 2004, **20**:475-9.
- Vitreschak AA, Rodionov DA, Mironov AA, Gelfand MS: **Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation.** *Nucleic Acids Research* 2002, **30**:3141-3151.
- Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS: **Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element.** *RNA* 2003, **9**:1084-1097.
- Heery DM, Dunican LK: **Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of Corynebacterium glutamicum: Identification of a mutation in the trp leader sequence.** *Applied and Environmental Microbiology* 1993, **59**:791-799.
- Lin C, Pradkar AS, Vining LC: **Regulation of an antranilate synthase gene in Streptomyces venezuelae by trp attenuator.** *Microbiology* 1998, **144**:1971-1980.
- Craster HL, Potter CA, Baumberg S: **End-product control of branched-chain amino acid biosynthesis genes in Streptomyces coelicolor A3(2): paradoxical relationships between DNA sequence and regulatory phenotype.** *Microbiology* 1999, **145**:2375-2384.

23. Konan KV, Yanofsky C: **Rho-dependent transcription termination in the *tna* operon of *Escherichia coli*: Roles of the boxA sequence and the *rut* site.** *Journal of Bacteriology* 2000, **182**:3981-3988.
24. Washio T, Sasayama J, Tomita M: **Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination.** *Nucleic Acids Research* 1998, **26**:5456-5463.
25. Unniraman S, Prakash R, Nagaraja V: **Conserved economics of transcription termination in eubacteria.** *Nucleic Acids Research* 2002, **30**:675-684.
26. Richardson JP: **Rho-dependent termination and ATPases in transcript termination.** *Biochimica et Biophysica Acta* 2002, **1577**:251-260.
27. Richardson JP: **Structural organization of transcription termination factor Rho.** *The Journal of Biological Chemistry* 1996, **271**:1251-1254.
28. Kaplan DL, O'Donne M: **Rho factor: transcription termination in four steps dispatch.** *Current Biology* 2003, **13**:R714-R716.
29. Gong F, Yanofsky C: **Rho's role in transcription attenuation in the *tna* operon of *E. coli*.** *Methods Enzymol* 2003, **371**:383-391.
30. Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ: **Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs.** *Chem Biol* 2004, **11**:1729-1741.
31. Batey RT, Gilbert SD, Montange RK: **Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine.** *Nature* 2004, **432**:411-415.
32. Koonin EV, Ilyina TV: **Computer-assisted dissection of rolling circle DNA replication.** *BioSystems* 1993, **30**:241-268.
33. Baytaluk MV, Gelfand MS, Mironov AA: **Exact mapping of prokaryotic gene starts.** *Briefings in Bioinformatics* 2002, **3**:181-194.
34. Gorbunov KYu, Mironov AA, Lyubetsky VA: **Search for conserved secondary structures of RNA.** *Mol Biol* 2003, **37**:850-860.
35. Lyubetsky VA, Seliverstov AV: **Selected algorithms related to finite groups.** *Information Processes* 2003, **3**:39-46. (in Russian)

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

