ISSN 0026-8933, Molecular Biology, 2010, Vol. 44, No. 1, pp. 128–139. © Pleiades Publishing, Inc., 2010.
Original Russian Text © K.V. Lopatovskaya, A.V. Seliverstov, V.A. Lyubetsky, 2010, published in Molekulyarnaya Biologiya, 2010, Vol. 44, No. 1, pp. 140–151.

BIOINFORMATICS

UDC 575.852

Attenuation Regulation of the Amino Acid and Aminoacyl-tRNA Biosynthesis Operons in Bacteria: A Comparative Genomic Analysis

K. V. Lopatovskaya, A. V. Seliverstov, and V. A. Lyubetsky

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

Received February 13, 2009
Accepted for publiation June 5, 2009

Abstract—A large-scale search for attenuation regulation in bacteria was performed using two original computer programs, which modeled the attenuation regulation and multiple alignment along a phylogenetic tree. The programs are available at http://lab6.iitp.ru. Candidate attenuations were predicted for many organisms belonging to α-, β-, γ-, and δ-proteobacteria, Actinobacteria, Bacteroidetes/Chlorobi, Firmicutes, and Thermotoga; in Cloroflexi, the corresponding sites were found upstream of *hisG, hisZ, hisS, pheA, pheST*, *trpEG, trpA, trpB, trpE, trpS, thrA, thrS, leuA, leuS, ilvB, ilvI, ilvA, ilvC, ilvD*, and *ilvG*. Searches were conducted across all bacterial genomes contained in GenBank, NCBI. Other bacterial taxa were not predicted to have attenuation. It was possible to assume, in some cases, that RNA triplexes play a substantial role in the formation of an active antiterminator and terminator or pseudoknots during termination. The attenuation regulation of *Lactobacillus lactis lysQ* was assumed to depend on the histidyl-tRNA concentration. Several types of attenuation regulation and the evolution of attenuation are discussed.

DOI: 10.1134/S0026893310010164

Key words: gene expression in bacteria, attenuation regulation, large-scale searches, attenuation prediction algorithms, tree-based multiple alignment algorithms

INTRODUCTION AND FORMULATION OF THE PROBLEM

This paper continues our previous publication [1] as regards both the method designed to identify potential attenuation structures and the large-scale search; in particular, we suggest new variants of attenuation regulation and examine the role of RNA triplexes and pseudoknots in the process. The title of this work continues that of the previous publication [1], which has included a detailed review of the relevant literature. Hence, the historical aspect of the problem is only considered in brief below.

Bacteria utilize various mechanisms to regulate the gene expression at the levels of transcription and translation, as well as posttranscriptional and posttranslational modification. The most comprehensive studies focus on the regulation based on protein–DNA interactions (repression and activation of transcription initiation), which has been considered in many studies, the regulation based on T-box formation [2–4] or riboswitches [5, 6], and the attenuation regulation, which has been described by Yanofsky and colleagues [7–9]. The references to the studies addressing the attenuation regulation and describing the relevant illustrations have also been given in the introduction to our previous publication [1]. Posttranscriptional and posttranslational modifications have been considered in many studies, of which [5, 10] are of particular interest. In many proteobacteria and actinobacteria, an attenuation structure regulates the *trp, his, leu*, and *ivl* operons, which code for amino acid synthesis enzymes and aminoacyl-tRNA synthases [11, 12]. The attenuation regulation depends on the concentration of the corresponding aminoacyl-tRNA and often involves several operons of one genome, namely, the operons coding for the enzymes producing the given amino acid and aminoacyl-tRNA synthases.

In this study, the term "attenuation regulation" is applied to the structure that is responsible for such regulation and includes the leader peptide gene with its regulatory codons and the associated alternative mRNA secondary structures. Some of these structures are known as terminating or sequestering structures and determine a premature termination of transcription of the structural gene or suppress its translation, while antiterminating or antisequestering structures allow RNA polymerase to continue transcription of the structural gene. The alternative depends on the rate at which the ribosome synthesizes the leader peptide or, more exactly, the rate at which the ribosome proceeds through regulatory codons. In turn, this rate depends on the concentration on the target substrate or a substrate-bound substance. Terminating and sequestering structures are conserved to a greater extent in all cases as compared with antiterminating and antisequestering structures.

The term "antiterminator" is hereafter applied to the hairpin that is rapidly degraded by the ribosome during its rapid progress along the leader peptide gene and plays a regulatory role. The term "terminator" refers to the hairpin that usually leads to a premature termination of transcription of the structural gene.

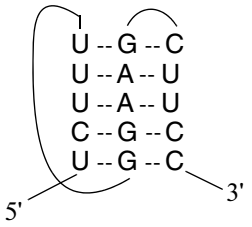
The most important type of the attenuation regulation is classical regulation, which occurs at the level of structural gene transcription, and whose terminating and antiterminating secondary structures are alternative in a certain sense of the word. The classical attenuation regulation includes the following cases. One case is Yanofsky's regulation, where the terminator and antiterminator are directly alternative to each other and there is a U-rich region (polyuracil tract) in the vicinity of the 3' end of the terminator. Another case is a chain of helices. The terminator and antiterminator are not alternative, but there is a chain usually consisting of four hairpins. Once formed, one hairpin, which acts as an antiterminator, prevents another (second, or coterminator) hairpin; then, a next (third, or coantiterminator) hairpin forms to prevent a terminator (fourth) hairpin; the mechanism also involves a U-rich tract located close to the 3' end of the terminator. In this case, we usually assume that the coterminator and/or coantiterminator are stabilized via the formation of RNA triplexes. A third case is an antiterminator ensemble. A conserved antiterminator is absent, and its role is played by an ensemble of hairpins, which each are alternative to a conserved terminator. As in the above cases, a polyuracil tract is involved, and the stability of certain hairpins is due to RNA triplex formation.

A U-rich region (polyuracil tract) is a sequence that usually consists of approximately seven nucleotides, mostly uracil residues. The other terms have been defined in [13].

A particular case of the classical attenuator regulation without a U-rich region is where the terminator and antiterminator are directly alternative to each other, but a U-rich region is lacking or contains two or three uracil residues at appreciable distances from each other. This case seems to occur actually, but it should be distinguished from the case described below, which is intermediate between the classical attenuator regulation and translational regulation.

A sequester-attenuation regulation involves all elements of the Yanofsky elements apart from a U-rich region, but occurs at the translational level. A hairpin similar to a terminator overlaps the ribosome-binding site.

A terminator ensemble is difficult to identify using a multiple sequence alignment, since each of the terminators is conserved to a low extent. In general, a multiple sequence alignment is also insufficient for distinguishing the variants of the regulation without a U-rich region (such cases are omitted below).



5' CCUGACUAG**UCUUU**CAGGCGAUGUGUG
CU**GGAAG**ACAUUCAGAU**CUUCC**AGUG 3'

Fig. 1. RNA triplex upstream of *E. coli hisG*. The third arm of the triplex is UCUUU; the helix arms are GGAAG and CUUCC.

RNA strands may form triplexes. A triplex includes three regions, of which one consists entirely of purines and the other two are, respectively, parallel and antiparallel to the first region (Fig. 1). The pair of the antiparallel regions form an RNA helix, wherein the Watson–Crick base pairs and a GU pair are considered to be complementary. Thus, two regions of a triplex form a standard helix, while the third (parallel) region is linked to the helix via Hoogsteen hydrogen bonds. This region is known as the third arm of the triplex; its nucleotides are separated with * from nucleotides of the helix in nucleotide sequences. The nucleotide triads possible for a triplex include C*GU, G*GC, G*GU, U*AU, A*AU, A*GC, C*GC, and certain others. Triads have several features. For instance, C*GC is only stable in a weakly acidic environment, where cytosine N3 is protonated. RNA triplexes have been considered in many publications, [14–19] being worthy of note here. When the third arm is at the 5' end of the helix, it is essential for triplex folding that a certain distance separate the third arm from the neighbor helix arm. We assumed that a third arm of l nt is at least $l + 6$ nt away from the neighbor arm. When the third arm is at the 3' end of the helix, they may be immediately adjacent to each other [18, 19].

As a nonclassical attenuation regulation, we considered the LEU1 and LEU regulations in addition to the above sequester-attenuation regulation in this work. Their structures are determined by a leader peptide gene with internal regulatory codons and the formation of pseudoknots shown in Figs. 2 and 3, respectively.[1] In our case, a pseudoknot is formed by two helices, one having an arm in the outer loop of the other. Of these helices, the one whose 5' end is closer to the 5' end of the original sequence is termed left, and the one whose 5' end is in the outer loop of the left helix is termed right. The pseudoknot itself is in the outer loop of the left helix. The helix that is closest to the pseudoknot and whose loop harbors the pseudoknot is termed the pseudoknot stem. Although

[1] Figures 3–37, their captions, and Table 2 are available at http://www.molecbio.com/downloads/2010/1/supp_lopat_rus.pdf

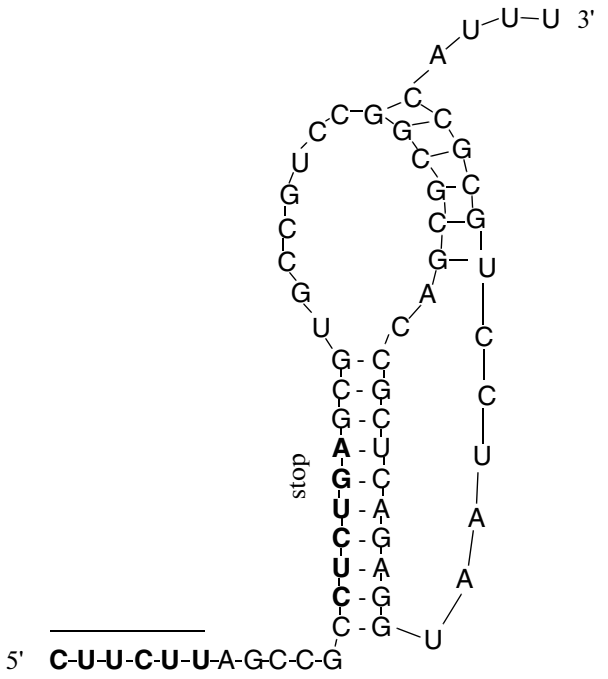


Fig. 2. LEU1 pseudoknot upstream of *D. shibae leuA*. The corresponding regulation involves a leader peptide gene with leucine codons. The left arm of the helix is in the vicinity of these codons. A stem is absent.

structurally similar to a certain extent, LEU and LEU1 pseudoknots substantially differ in size, nucleotide composition, and the nucleotide positions relative to the structural and leader peptide genes. In particular, LEU structures usually have a stem, while LEU1 structures do not.

The mechanism of the LEU regulation is overlapping the ribosome-binding site. When a ribosome proceeds through the leader peptide gene rapidly, a stem forms; when a ribosome is stalled on the regulatory codons, it prevents the formation of a stem (Fig. 3). This mechanism has been considered in detail in our earlier study [2]. The origin of the LEU element might involve its horizontal transfer from an ancestor of Bifidobacteriales to an ancestor of Actinomycetales, since *Bitiaobacteria longum* still preserves a transposase open reading frame harboring a pseudoknot, while this open reading frame was evolutionarily transformed into a regulatory element upstream of *leuA* in Actinomycetales [2]. The mechanism and evolution of the LEU1 regulation are less clear; they are considered in Results and Discussion in connection with the regulation observed in α-proteobacteria.

In this study, we checked whether the attenuation regulation of *hisG, hisZ, hisS, pheA, pheST, trpE, trpEG, trpA, trpB, trpS, thrA, thrS, leuA, leuS, ilvB, ilvI, ilvA, ilvC, ilvD*, and *ilvG* expression is found in the total diversity of sequenced bacterial genomes available from NCBI GenBank [20, 21]. As a result, numerous new types of the attenuation regulation were predicted in several taxa, while our methods did not detect it upstream of the above genes in other taxa.

The above protein-coding genes were chosen by the systematic presence of a leader peptide gene with regulatory codons for the amino acids that are biochemically associated with the downstream structural gene.

The functions of the genes are as follows. The *hisG* and *hisZ* genes code for subunits of ATP-phosphoribosyltransferase [EC 2.4.2.17], which catalyzes the formation of N'-5'-phosphoribosyl-ATP of phosphoribosyl pyrophosphate. HisG and HisZ form a heteromultimer. The *hisS* gene codes for histidyl-tRNA synthetase. The *pheA* gene codes for prephenate dehydratase [EC 4.2.1.51], which is involved in aromatic amino acid synthesis. The *pheST* operon codes for the α and β subunits of phenylalanyl-tRNA synthetase; the subunits form the $\alpha_2\beta_2$ heterotetramer and bind two magnesium ions. The *trpA* and *trpB* genes code for tryptophan synthase subunits, while the *trpE* gene or the *trpEG* operon codes for anthranilate synthase subunits, which often occur as one protein. The *trp*S gene codes for tryptophanyl-tRNA synthetase. The *thrA* gene codes for bifuntional aspartokinase/homoserine dehydrogenase, while *thrS* codes for threonyl-tRNA synthase. The *leuA* gene codes for 2-isopropylmalate synthase. In many γ-proteobacteria, *leu*A belongs to the *leuABCD* operon, whose genes are involved in leucine synthesis. In other species, *leuA* is not included in a polycistronic operon. The *leuS* gene codes for leucyl-tRNA synthetase. The *ilvD* gene codes for dihydroxyacid dehydratase, which dehydrates 2,3-dihydroxy-3-methylbutanoate to yield 3-methyl-2-oxybutanoate in valine and isoleucine syntheses. Many γ-proteobacteria have the *ilvGMEDA* operon, which codes four out of the five enzymes of the isoleucine and valine biosynthesis pathways. The *ilvB* gene codes for the large subunit of acetolactate synthase and often occurs as part of the *ilvBNC* or *ilvBHC* operon, where *ilvN* and *ilvH* code for the small acetolactate synthase subunit, and *ilvC* does for ketol-acid reductoisomerase. The *ilvA* gene codes for threonine dehydratase.

Compare the method used in this study and considered below with the method described earlier [1]. The method [1] employs the LLLM program, which utilizes special criteria to consequently search the leader region of a structural protein-coding gene (usually for an enzyme involved in biosynthesis of a certain amino acid or for an aminoacyl-tRNA synthase) for possible positions of the leader peptide gene. Then, the program identifies the possible variants of the U-rich region and terminator hairpin for each position of the leader peptide gene and finds the possible antiterminator hairpins for each of the variants identified. Next, the resulting leader regions, which potentially harbor the structures involved in the classical attenuation regulation, are aligned using a standard local alignment program. If the local multiple nucleotide sequence alignment obtained for the leader gene

regions is of good quality, it is considered to predict the classical attenuation regulation upstream of the corresponding gene in the corresponding species. Thus, the method [1] consists in selecting the leader regions by the LLLM program and searching for a good multiple sequence alignment where the elements of the classical attenuation regulation occurring in the selected regions are well aligned with each other. It is clear that the LLLM program is purely combinatorial and does not describes the time course (that is, the process) of the classical attenuation regulation.

We have earlier proposed an attenuation model [13]; the model is implemented in the rapid RNA-model program, which is described in detail in [22] and available at http://lab6.iitp.ru/. The model is similar to the LLLM program in being applied to individual leader regions, but has essentially different procedures. Substantial changes were made to the model and the RNAmodel program for the purposes of this study (see Method). Thus, we used the RNAmodel program in place of the LLLM program in this work. When a distinct (informative) pattern was obtained, an additional search for the regulation was performed using the pattern as a reference.

A similar approach was used to identify the LEU and LEU1 regulations in individual leader regions, but the model and programs described earlier [13, 22] were replaced with their adapted variants, which were termed LEUmodel and LEU1model, respectively. These programs are briefly described in Method and are as of yet unavailable at our web site.

After individual leader regions had been examined using the RNAmodel, LEUmodel, and LEU1model programs, we constructed a multiple sequence alignment of the selected leader regions potentially containing the attenuation regulation structure, as in the earlier study. In contrast to [1], a different program was employed in multiple sequence alignment. Namely, we used an original program that constructed a multiple sequence alignment along a phylogenetic tree, which might be polytomic [23]. The program is briefly described below (see Method) and is available at http://lab6.iitp.ru.

To avoid underprediction, the LEU and LEU1 regulations were searched against a reference along the total genomes of several bacteria, rather than in the leader re the above genes. The search did not yield new results as compared with our earlier study [2], when a fragment homologous to the LEU element was found within the transposase gene of *B. longum*.

METHOD

Bacterial genomes were sought using the BLAST program [23]. Phylogenetic trees were constructed by the neighbor-joining algorithm with the use of the CLUSTALW program [24]. Our method employed the first and second algorithms, which are considered below and are available at http://lab6.iitp.ru. The first algorithm is an improvement of our earlier algorithm [13]; the second algorithm is described for the first time, except for a brief annotation [23].

The first algorithm. The classical attenuation regulation model [13] is based on a simultaneous description of ribosome progress along the leader peptide gene, RNA polymerase progress up to the end of the U-rich region and the start of the structural gene, and the formation of a secondary structure in the region between the 3' end of the ribosome and the 5' end of polymerase (the region is termed the "window"), which is followed by a short region where the DNA mRNA duplex is bound in a channel formed by the polymerase subunits. The model describes the probability $p(c)$ of premature polymerase termination as a function of the aminoacyl-tRNA concentration c; the probability is calculated taking the effect of the concentration c on the rate of ribosome progress through regulatory codons into account. For a given nucleotide sequence and a given c value, the model constructs a chain of typical, in a certain sense, secondary structures in each window. The chain depends on the kinetics of secondary structures and the position of the window resulting from the movements of the ribosome and polymerase. The chain together with the consecutive positions of the ribosome and polymerase is termed the modeling trajectory.

A key improvement of our earlier model [13] is that we no longer isolate a pause hairpin, antiterminator, and terminator with their specific effects on polymerase, but rather describe how every helix of the current secondary structure in a window uniformly interacts with all other helices, the ribosome, and polymerase, slowing down the polymerase progress to a greater or lesser extent or causing its premature termination (as a limiting case of such deceleration), depending on the concentration c. This idea agrees with our observations of the model [13]: the attenuation regulation is accompanied by the formation of an ensemble of terminators or, more often, an ensemble of antiterminators; an antiterminator may have arms reaching the U-rich region and may form a pseudoknot with a terminator; etc. The situation is further complicated by the fact that, when a leader peptide gene with internal regulatory codons is present, the U-rich region may include only one or two uracil residues at an appreciable distance from each other (a weak polyuracil tract), or there may be no such region at all. Such a situation suggests the classical attenuation regulation in some cases or, in other cases, is indicative of another regulation involving the leader peptide gene and its regulatory codons. The latter case includes the sequester-attenuation, LEU, and LEU1 regulations. It should be noted that a change in exactly one position of the leader sequence is capable of abolishing the attenuation regulation, which was distinct before [7]. Our model reproduces this phenomenon.

An analysis of the modeling trajectories in our model makes it possible to ascribe a terminator role to certain hairpins and an antiterminator to some others. This makes it possible to use the second algorithm (see below) and to construct a local multiple sequence alignment of potential regulatory regions that is adequate to the regulation, if any. The first algorithm is compared with the LLLM algorithm [1] in Introduction.

In addition, the following changes were made to the classical attenuation regulation model compared with the earlier one [13]. Pseudoknots are allowed for all secondary structures in every current window. Energy of the secondary structure is calculated using an improved variant of the earlier method [25], and the temperature typical of the habitat of the given bacterium is taken into account using an improved variant of another method [26]. In addition, the model allows for the formation of RNA triplexes, which stabilize certain helices of the secondary structure that are critical for the regulation by reducing their energy. The corresponding RNAmodel program is available at http://lab6.iitp.ru. In the mode allowing for pseudoknots, the dependence $p(c)$ was qualitatively preserved for all sequences examined in the mode with pseudoknots prohibited in [13] and was even improved in many cases.

We performed a large-scale testing of the model and, therefore, the RNAmodel program. That is, the output of the program was compared with the available results of bioinformatics or experimental prediction of the classical attenuation regulation in bacteria, including the cases where a lack of such regulation was predicted. The comparison yielded the following result. When the regulation had been predicted, the model reported a strengthening concentration dependence of probability $p(c)$ in a certain concentration range with a high significance. Vice versa, when the absence of the regulation had been predicted, $p(c)$ obtained with the model was constant or decreasing. The comparison results are available at http://lab6.iitp.ru/rnamodel (file Mass). Modeling was performed with the same universal parameters, which are specified in [21] and on the above web page.

In the case of the LEU regulation, we developed a model and the LEUmodel program, which determines the probability $q(c)$ of the noninitiation of translation of a structural gene as a function of the leucyl-tRNA concentration. This probability is determined as the mRNA life portion during which the putative ribosome-binding site is overlapped by any helix. The greatest contribution to this overlap is made by the stem of the LEU pseudoknot shown in Fig. 3. In the cases where the LEU regulation had been predicted, the model reported a decreasing $q(c)$ dependence; when such a regulation had not been predicted, the $q(c)$ dependence was essentially other than decreasing.

In the case of the LEU1 regulation, we developed a variant of the above model and the LEU1model program. The program determines the probability $q_1(c)$ of a regulatory event as a function of the leucyl-tRNA concentration. The probability is determined as the mRNA life portion during which there exist hypohelices of the two fixed helices that are conserved to the greatest extent and form the LEU1 pseudoknot (Fig. 2).

The second algorithm. Below is a brief description of the algorithm and the LmalTree program, which constructs a multiple sequence alignment along a phylogenetic tree. The program was used to obtain local alignments of sequences potentially harboring the relevant regulatory structures (the sequences were selected using the first algorithm) [22]. As the RNAmodel program, the LmalTree program is available at http://lab6.iitp.ru.

First, we consider the algorithm applied to a binary phylogenetic tree of species. Each branch of the species tree is ascribed with one sequence of nucleotide frequency distributions (hereafter referred to as a sequence of distributions); i.e., the members of a sequence are vectors of a length of four. A vector shows the frequencies of the four nucleotides in the following order: A, C, G, and T. Thus, every position of a sequence of distributions contains a vector, which is the nucleotide frequency distribution characteristic of the given position. Leaves of the tree are ascribed with nucleotide sequences; they are considered as the sequences of distributions that consist of vectors wherein one frequency is unity and the other frequencies are zero.

First, the algorithm processes the branches of a tree from leaves to the root (a forward run of the algorithm), consecutively ascribing the sequences of distributions to them. When two distributions have been ascribed to two "sons" of a particular branch, the sequences are aligned and, then, their "father" is ascribed with a half-sum of the aligned distributions of the sons. The alignment is constructed using a standard pairwise alignment algorithm, but the alignment quality functional (score) is calculated in a specific way. While fixed bonuses/penalties for matching/mismatching nucleotides are used in aligning nucleotide sequences, our algorithm calculates the score a_j for each gap-free j-th position of the alignment; i.e., a_j may be obtained as a scalar square difference between two vectors: $a_j = 1 - \sum_{i=1}^{4} w_i(x_i - y_i)^2$, where w_i are the nucleotide weights, whose sum is unity. If a gap is opened in a certain position of one of the sequences of distributions during alignment, the gap corresponds to a zero distribution. In this case, a_j is replaced by a usual penalty function with the penalty a_j' per position decreasing with an increasing length of the series of consecutive gaps. The final pairwise alignment score,

Potential attenuation regulation in bacteria

Types and classes	Genes						
α-proteobacteria	*ilvB, I*	*trpE*	*hisS*	*pheST*	*thrA*	*leuA*	*leuA*
β-proteobacteria	*ilvB*	*trpE*		*pheA*	*thrS*	*leuA*	*leuA, ilvB*
γ-proteobacteria	*ilvB, G*	*trpE*	*hisG*	*pheA, S*	*thrA*	*leuA*	
δ-proteobacteria	*ilvB*	*trpS*			*thrA, S*	*leuA*	
Actinobacteria	*ilvB, I, D*	*trpE, S, BE, BA*					*leuA*
Bacteroidetes/Chlorobi	*ilvD*	*trpE*	*hisG*				
Firmicutes	*ilvD, lysQ*	*trpB*	*hisZ*				
Thermotogae		*trpE*	*hisS*				
Chloroflexi	*ilvD*						

Note: Empty cells indicate that the regulation is absent. All but the last column show the genes that are subject to the classical attenuation regulation; the last column shows the genes subject to the LEU1 regulation in proteobacteria and the LEU regulation in Actinobacteria.

whose maximum is sought, is determined as a sum of the a_j or $a_j^{'}$ values over all positions of the alignment.

After constructing a pairwise alignment, a sequence of distributions is obtained for the father by calculating Z for each of its position as the half-sum of the distributions X and Y occurring in the same position in its aligned sons; i.e., $Z = \frac{1}{2}(X + Y)$.

If a sequence of distributions has already been constructed for the root of a tree, a reverse run of the algorithm is performed. First, the gaps opened in the sequences corresponding to root-neighboring branches are extended along the tree up to its leaves. Then, the same procedure is performed with gaps opened during alignments at the previous tree level (the third level, counting from the root) and is repeated consecutively up to the level located immediately at the top of tree leaves.

Generally speaking, the sequences with numerous gaps that have been constructed for tree leaves by this procedure represent the desirable multiple alignment and the result produced by the algorithm for the given binary tree.

However, a known phylogenetic tree of species is not binary in many cases, but rather contains at least one polytomic branch (a polytomic tree). In this case, the algorithm resolves all nonbinary (polytomic) branches by adding intermediate branches to the original nonbinary tree. This yields all variants of binary trees that agree with the original nonbinary tree as far as the x—progeny y relationship is concerned. The series of binary trees should be generated without repetitive variants because the variant number is already great, since each polytomic branch with n sons allows $\frac{(2n-3)!}{2^{n-2}(n-2)!}$ topologically different solutions. When there are several polytomies, each of them is resolved independently, and the total number of trees is the product of the numbers of variants over all nonbinary branches.

The multiple sequence alignments obtained for different binary solutions of the polytomic tree are compared by calculating the parameter $(N_a + N_s)b + \sum_{i=1}^{N_s}(b+s)(l_i - 1) + N_b c$, where N_a is the number of absolutely conserved single (i.e., having exactly one symbol in the column) positions in the alignment; N_s is the number of absolutely conserved continuous regions of at least 2 nt (l_i is the length of the i-th region); N_b is the number of almost absolutely conserved (i.e., having only one different symbol in the column) positions in the alignment; b, c, and s are the bonuses, which are the parameters of the algorithm.

RESULTS AND DISCUSSION

Our method revealed potential attenuation structures in species belonging to α-, β-, γ-, and δ-proteobacteria, including the proteobacterium *Magnetococcus* sp. MC-1; Actinobacteria; Bacteroidetes/Chlorobi; Firmicutes; Thermotogae; and Cloroflexi (table). The method did not detect an attenuation structure in any species of Chlamydiae, Cyanobacteria, Mollicutes, ε-proteobacteria, and Spirochaetales. The structures were not found in algal chloroplasts, which have amino acid synthesis genes.

The attenuation regulation is the most common in proteobacteria and Actinobacteria. In individual cases, it was predicted for representatives of other taxonomic groups: Firmicutes, Bacteroidetes/Chlorobi, Thermotogae, and Cloroflexi. The frequencies of such regulation of different genes substantially vary.

As is evident from table, the greatest number of species and the greatest diversity of amino acid and aminoacyl-tRNA biosynthesis genes subject to the classical attenuation regulation were characteristic of γ-proteobacteria. A nonclassical attenuation regulation of *leuA* was observed for Actinobacteria and α-proteobacteria.

The results obtained for individual amino acids are considered below.

Phenylalanine

The classical attenuation structure depending on phenylalanyl-tRNA was observed exclusively in α-, β-, and γ-proteobacteria, regulating transcription of the *pheA* phenylalanine synthesis gene and the *pheST* operon. The regulation of the *pheST* operon was predicted for a few α-proteobacteria of the order Rhodobacterales and the proteobacterium *Magnetococcus* sp. MC-1. The regulation was predicted for *phe*A in the genus *Bordetella* (the class β-proteobacteria) and for *pheA* and *pheS* for the majority of γ-proteobacteria (Figs. 4, 5)[2] In *Psychromonas* sp. CNPT3, the attenuation regulation was observed only upstream of *pheST*. In general, the regulation of *pheST* was less common than that of *pheA*. It is seen from the tree that this regulation is mosaically distributed among α-, β-, and γ-proteobacteria, occurring in single representatives of the majority of taxonomic groups. The regulation was found upstream of both *pheA* and *pheST* in representatives of the orders Enterobacteriales and Alteromonadales and only upstream of *pheA* in representatives of the orders Aeromonadales and Vibrionales. It is seen from the tree that the regulation is mosaically distributed among α-, β-, and γ-proteobacteria; i.e., the distribution is nonuniform in various taxonomic groups.

A tree of regulatory regions was constructed on the basis of the corresponding alignment (Fig. 6). In γ-proteobacteria, the structures upstream of *pheST* formed a clade with the only exception of *Psychromonas* sp. CNPT3. This makes it possible to assume that the attenuation regulation of *pheST* appeared in a common ancestor of γ-proteobacteria and evolved independently of the *pheA* regulation. The regulatory region upstream of *pheST* in *Psychromonas* sp. CNPT3 substantially differs from that in other species. This indicates that the attenuation regulation of this region arose in this species independently of the *pheST* regulation of most other species. It is of interest that the *Psychromonas* sp. CNTP3 region is close to the regulatory region upstream of *pheA* of *Alteromonas macleodii*, which belongs to the same order Alteromonadales. Horizontal gene transfer or duplication with subsequent loss of the corresponding regulatory structures can be assumed in this case.

Threonine and Isoleucine

The classical attenuation structure associated with threonine and isoleucine was found only in proteobacteria. In α-proteobacteria, this structure regulates only *thrA*, which is involved in amino acid synthesis, and only in *Rhodobacterales bacterium* (Fig. 7a). In γ-proteobacteria, the structure regulates only the same *thrA* gene, but that in many species (Fig. 8). In β-proteobacteria, the structure regulates only *thrS*, which codes for threonyl-tRNA synthetase, in several species (Fig. 9).

In δ-proteobacteria, the structure was predicted for *thrA* of *Myxococcus xanthus*. A weak prediction was obtained for two *thrS* paralogs in *Bdellovibrio bacteriovorus* and one of the two paralogs in *Myxococcus xanthus* and *Stigmatella aurantiaca* (Fig. 7b). It should be noted that, according to our model [13], an antiterminator forms at a low threonine concentration in *Myxococcus xanthus*, and this antiterminator prevents the formation of a terminator hairpin and is close to the U-rich region because of the greater arm lengths. At moderate concentrations, the ribosome partly disrupts the antiterminator, but the terminator may still form. At higher threonine concentrations, the ribosome moves up to the stop codon of the leader peptide gene and completely disrupts the antiterminator, allowing the formation of a short helix corresponding to the classical terminator. Such a phenomenon was also observed in other modeling cases. Since several regulations were additionally found upstream nonorthologous genes in δ-proteobacteria, it is possible to assume horizontal transfer of the regulatory structure without a transfer of the corresponding genes (Figs. 10, 11). Note that horizontal transfer can be assumed for several paralogs corresponding to these genes but lacking this regulation, e.g., for *thrS* of *Anaeromyxobacter dehalogenans* and *thrA* of *Desulfovibrio desulfuricans* and *Desulfatibacillum alkenivorans*.

Pairs of *thrS* paralogs that lack the classical attenuation regulation were observed in the δ-proteobacteria *Syntrophobacter fumaroxidans* and *Anaeromyxobacter dehalogenans*.

Tryptophan

In Actinobacteria, the classical attenuation structure was predicted for the operons that harbor *thrE* in *Streptomyces* and three of the four *Corynebacterium* species (*C. diphtheriae, C. glutamicum,* and *C. efficiens*) (Fig. 12a). A search by reference indicated that this regulation occurs upstream of *thrS*. Our method did not predict the regulation for *C. jeikeium*.

The *C. diphtheriae* regulatory regions align upstream of the two *thrB* paralogs. Of these, one belongs to the *thrBA* operon, which codes for tryptophan synthase, and the other belongs to the *thrBEGDC* operon, which codes for anthranylate synthase. However, the alignment upstream of *thrBA* is far worse in quality. A modeling also testified to a lack of the regulation upstream of *thrBA*. The regulation was not found upstream of *thrBA* in other actinobacteria. It is advantageous to regulate the genes for anthranylate synthase subunits (which are involved in the first step of the tryptophan biosynthesis pathway), while the

[2] Figures 3–37, their captions, and Table 2 are available at http://www.molecbio.com/downloads/2010/1/supp_lopat_rus.pdf

regulation is unnecessary for the genes for tryptophan synthase subunits (the last step of the same pathway). This circumstance may explain the substantial divergence of the attenuation structure upstream of *thrBA*. It is possible that *Corynebacterium diphtheriae* experienced a series of chromosome rearrangements that placed *thrB* between the regulatory region and structural part of *thrE* and were followed by a duplication of *tthrB* with its regulatory region. As a result, the *thrB* paralog together with its regulatory region was spatially associated with *thrA*. Then, mutations accumulated in the regulatory region upstream of *thrBA*.

The branches leading to *thrS* of *Streptomyces avermitilis* and *thrE* in *Streptomyces* spp. are close on a tree of regulatory regions (Fig. 13). Hence, it is possible to assume that the *S. avermitilis* regulation upstream of *thrS* originates from the more ancient regulation upstream of *thrE*. The regulation upstream of *thrS* in *Nocardia farcinica* and *Saccharopolyspora erythraea* is of an ancient origin (Fig. 13).

Many cases of the classical attenuation regulation of *thrE* were predicted in α-, β-, and γ-proteobacteria, the order Bacteroidetes, and two representatives of *Thermotoga* spp. (Thermotogae) (Figs. 14–17). The regulation upstream of *thrS* with the corresponding ensemble of helices was predicted in δ-proteobacteria (Fig. 18). The regulation upstream of *thrB* was predicted for two representatives of *Bacillus* spp. (Firmicutes) (Fig. 12b).

Only one (downstream) of the two *thrE* paralogs has the regulation in *Vibrio fischeri*. The regulatory region of *thrE* was duplicated in the γ-proteobacteria *Pseudoalteromonas tunicata* and *Alteromonadales bacterium* (Fig. 16).

All of the above cases correspond to Yanofsky's regulation with the only exception of *thrS1* of δ-proteobacteria.

Histidine

The histidine-related classical attenuation structure was predicted to regulate transcription of *hisS* in α-proteobacteria, *hisG* in γ-proteobacteria, *hisZ* in Firmicutes, *hisG* in Bacteroidetes, and *his*S in Thermotogae (Figs. 19–22). The regulation is based on a chain of helices and involves triplexes in γ-proteobacteria and follows Yanofsky's pattern in Bacteroidetes, Thermotogae, and the genus *Listeria*.

A triplex involved in the formation of a coterminator according to our model was found upstream of *his*G in many γ-proteobacteria, including representatives of the families Enterobacteriales, Pasteurellales, Vibrionales, Alteromonadales, and Aeromonadales. The third arm of the triplex contains many uracil residues, which ensure its high stability regardless of the acidity of the cytoplasm. The triplex consists of Py-Pu-Py triads in the majority of cases. However, the combined CUGU*GAGG-CCUC triplex, which contains Py-Pu-Py and Pu-Pu-Py triads, forms in *Alteromonadales bacterium* and *Pseudoalteromonas haloplanktis* (Fig. 20). Again, the third arm of the triplex is indicated with an asterisk. The regulatory regions shown in Figs. 20 and 22 align well, suggesting their high conservation among many γ-proteobacteria and Bacteroidetes.

A weak cytidyl-guanyl Py-Pu-Py triplex is upstream of *hixZ* and is possibly involved in the formation of a coterminator in several bacilli (*Bacillus cereus, B. thuringiensis, B. anthracis*, and *B. weihenstephanensis*). In *Clostridium difficile*, a coterminator upstream of *hisZ* involves the AAG*AAG-CUU Pu-Pu-Py triplex (Fig. 21).

The classical attenuation regulation with histidine regulatory codons and a chain of helices wherein a coterminator is maintained by the AGA*AGA-UCU Pu-Pu-Py triplex was observed upstream of the *lysQ* permease gene in *Lactococcus lactis* (Firmicutes) (Fig. 23). It is possible that this permease changed its specificity to act as a histidine transporter only in *Lactococcus lactis* among all Firmicutes, since the regulation depends on the histidine concentration, while the orthologous genes of other Firmicutes lack even a leader peptide gene in their upstream regions.

While a chain of helices with triplexes was observed upstream of *his*Z in *Bacillus* and *Clostridium*, Yanofsky's classical attenuation regulation was predicted for species of the phylogenetically close genus *Listeria*.

Triplexes were found only in small taxonomic groups of Firmicutes, suggesting their recent evolutionary origin. Since the stability of several triplexes depends on the acidity or ionic strength of the cytoplasm, it is possible to assume that the role of the triplexes in the regulation is determined by intricate feedbacks, which coordinate the response to a certain concentration of the relevant amino acid with the acidity of the cytoplasm.

A modeling that did not allow for triplex formation predicted a low efficiency of the regulation, while a modeling allowing for triplexes showed that the frequency of premature transcription termination substantially changes with the concentration of the corresponding aminoacyl-tRNA. Thus, the role of triplexes in the regulation in γ-proteobacteria was supported by both their conservation and the modeling results. A modeling of the leader region upstream of *hisG* in γ-proteobacteria similarly showed that, when triplexes are not allowed, the antiterminator RNA structures are not stable enough to ensure the efficient regulation. In the case of *Clostridium difficile* and *hisZ*, a histidyl-tRNA concentration dependence of structural gene expression became distinct enough to suggest the regulation only when triplexes were allowed and, consequently, coterminator energy increased.

Branched Amino Acids

A classical attenuation structure with an ensemble of antiterminators was predicted for *ilvB* of most acti-

nobacteria of the order Actinomycetales, except for *Propionibacterium acnes, Tropheryma whipplei, Clavibacter michiganensis, Leifsonia xyli, Janibacter*, and *Salinispora arenicola*. Apart from Actinomycetales, the regulation was predicted for *Bifidobacterium adolescentis*. However, our method did not predict such a regulation for *Bifidobacterium longum* of the same genus and *Rubrobacter xylanophilus*, since these bacteria lacked even the leader peptide gene. The majority of Actinomycetales have a conserved transcription terminator, while a conserved antiterminator is lacking, and its role is played by an ensemble of several helices and pseudoknots (Fig. 24).

Six *liv*B paralogs have been annotated for the actinobacterium *Rhodococcus jostii*, and two of them (*ilvB4* and *ilvB5*) have a leader peptide gene and a pair of long alternative hairpins in the upstream region. A helix, which presumably acts as a terminator, is followed by a U-rich region upstream of *ilvB5*, but not of *ilvB4*. The classical attenuation regulation without a U-rich region is possible here; it cannot also be excluded that there is no regulation in some of these cases. Yanofsky's classical attenuation regulation of another paralog is degraded, but some of its elements are still preserved. The regulatory regions of *ilvB4* and *ilvB5* substantially differ from each other (Fig. 24).

An ensemble of helices similarly acts as an antiterminator of *ilvB* in the majority of α-proteobacteria (Fig. 25). In *Rhodobacterales bacterium*, the regulation was observed for both of the *ilv*B paralogs shown in Fig. 25.

A transcriptional regulation of *ilvB* seems to be absent in β-proteobacteria of the genus *Bordetella*. When the initiator codon indicated for the structural gene in the NCBI annotation was replaced with the codon that aligns with the initiator codons of *B. bronchiseptica* and *B. parapertussis*, the translational regulation (i.e., the sequester-attenuation regulation) was predicted. Its elements are shown in Fig. 26a.

The *leuABCD* operon and the *ilvA, ilvC, ilvG*, and *ilvB* genes, which often belong to various operons, are subject to Yanofsky's classical attenuation regulation in many γ-proteobacteria (Fig. 27). Some of these cases have been considered earlier [1].

A tree of the regulatory regions found upstream of the *ilv* genes in γ-proteobacteria showed that the regulatory regions upstream of *ilvA, ilvC, ilvB*, or *ilvC* and *ilvB* taken together form four clades. The regulation upstream of *ilvG* is the most ancient, originating in a common ancestor of γ-proteobacteria (Fig. 28).

In the δ-proteobacterium *Stigmatella aurantiaca*, the upstream region of *ilvB* contains a potential structure that is similar to the classical attenuation regulation structure. The structure includes a leader peptide gene and has several specifics, lacking a U-rich region, etc. (Fig. 26c). A similar situation was observed for representatives of the orders Desulfobacterales, Desulfuromonadales, and Syntrophobacterales (δ-proteobacteria). However, the corresponding regulatory regions are not conserved, and the regulation is not supported by modeling (Fig. 29).

The classical attenuation regulation was predicted for *ilvD* in one actinobacterium, *Corynebacterium efficiens* (Fig. 30a).

In the genera *Staphylococcus* and *Listeria* (Firmicutes), the regulatory region of *ilvD* displays a chain of helices, which includes four conserved helices and forms a triplex of Py-Pu-Py triads upon the formation of a coantiterminator. An alternative antiterminator, which is longer and does not require triplexes, is also possible in this case (Fig. 31). A modeling of the *ilvD* regulation in *Staphylococcus* and *Listeria* (Fig. 32) revealed a situation similar to that with *hisG*, indicating the necessity to take energy of the triplex involved in the antiterminator into account. This finding, along with conservation, demonstrates again that triplexes should be considered when calculating energy for an RNA secondary structure. *Geobacillus thermodenitrificans* displayed the Yanofsky's classical regulation. The same regulation was observed for representatives of Bacteroidetes/Chlorobi and for *Herpetosiphon aurantiacus* (Chloroflexi) (Figs. 30b, 30c).

Yanofsky's classical attenuation regulation of *leuA* was predicted for a few α- and β-proteobacteria. Among these, *Acidovorax* sp. JS42 is unique in that its two *leuA* paralogs form an operon with a common regulation. As for β-proteobacteria, the classical attenuation structure was observed in representatives of Burkholderiales (*Burkholderia xenovorans, Burkholderia phytofirmans, Burkholderia pseudomallei, Delftia acidovorans*, *Polaromonas* sp., *Comamonas testosteroni, Methylibium petroleiphilum*, and *Acidovorax* sp.) (Figs. 33a, 33b).

The γ-proteobacteria Enterobacteriales, Pasteurellaceae, *Shewanella* spp., Vibrionaceae, and Xanthomonadaceae have a structure for Yanofsky's classical attenuation regulation, including a conserved terminator of a great length. Some of these species have been considered earlier [1]. The *leuA* gene lacks paralogs and is included in the extended *leuABCD* operon in these γ-proteobacteria except for Xanthomonadaceae. In Xanthomonadaceae, *leuA* is included in the *ilvCGM-tdcB-leuA* operon. Thus, the regulatory region is immediately of *leuA* in one case and is far away in the other, so that the *ilv* genes and *leuA* are regulated together. Similar situations were observed in several α-proteobacteria. The *leuA* gene similarly lacks paralogs in Pseudomonadaceae, but it is not included in a polycistronic operon, nor does it have a leader peptide gene (Fig. 34).

Yanofsky's classical attenuation regulation upstream of *leuA* and *ilvB* was predicted for a few δ-proteobacteria. In particular, *leuA* is regulated in *Desulfotalea psychrophila, Stigmatella aurantiaca, Syntrophobacter fumaroxidans, Syntrophus aciditrophicus*, and *Plesiocystis pacifica* (Fig. 33c). A tree of spe-

cies showed a mosaic phylogenetic distribution of the regulation of this gene in δ-proteobacteria.

A search by reference suggested Yanofsky's classical attenuation regulation upstream of *leuS* for *S. avermitilis* and *S. coelicolor* (Actinobacteria). However, a modeling did not confirm this regulation of the aminoacy-tRNA synthetase gene.

LEU regulation of *leuA* in Actinobacteria. We substantially extended the list of LEU regulations [2]; the regulation was found upstream of *leuA* in the majority of Actinomycetales (Fig. 35). The LEU regulation was observed in many actinobacteria of the order Actinomycetales with the exception of two families of Propionibacterineae (*Propionibacterium acnes* and *Nocardioides* sp. JS614) and the family Cellulomonadaceae (*Tropheryma whipplei*). The regulation was not found in the other orders of Actinobacteria, including Rubrobacterales (*Rubrobacter xylanophilus*), Coriobacteriales (*Atopobium minutum*), and Bifidobacteriales (*Bifidobacterium longum* and *B. adolescentis*).

It is possible to assume that the LEU regulation upstream of *leuA* arose in the last common ancestor of Actinomycetales and then disappeared in a common ancestor of Propionibacterineae, although the *leuA* gene itself is well preserved in this taxon. The *leuA* gene is lacking in *Tropheryma whipplei* (Cellulomonadaceae, Actinomycetales).

Craster et al. [24] have shown that a substitution of the leucine codons in the putative leader peptide gene upstream of *leuA* in *Streptomyces coelicolor* does not affect the intensity of transcription. This makes it possible to assume that the LEU element, which is highly conserved among Actinomycetales, is involved in the expression regulation at the level of translation initiation. This assumption is supported by the finding that the 3' arm of the stem of the LEU element overlaps the ribosome-binding site and; therefore, the stem prevents translation initiation of *leuA*. The 5' arm of the stem is within the leader peptide gene and is close to the leucine codons; consequently, the stability of the stem depends on the rate of *leuA* translation.

LEU1 regulation of *leu*A in proteobacteria. Many α-proteobacteria and several β-proteobacteria (Burkholderiales) have a leader peptide gene upstream of *leuA* that contains a sequence of leucine codons. In place of an RNA secondary structure characteristic of the classical attenuation regulation or the LEU regulation, this region forms a pseudoknot, which was termed the LEU1 pseudoknot (Fig. 2). The LEU1 pseudoknot along with a leader peptide gene and regulatory (in this case, leucine) codons were termed the LEU1 regulation (Figs. 36, 37).

The LEU1 regulation upstream of *leuA* was found in α-proteobacteria (Rhizobiales: *Agrobacterium tumefaciens, Aurantimonas* sp. SI85-9A1, *Brucella* spp., *Fulvimarina pelagi, Mesorhizobium* spp., *Rhizobium* spp., and *Sinorhizobium* spp.; Rhodospirillales: *Magnetospirillum* spp.; and Rhodobacterales: *Dinoroseobacter shibae, Jannaschia* sp. CCS1, *Loktanella vestfoldensis, Oceanicola* spp., *Rhodobacterales bacterium* HTCC2654, *Rhodobacter* spp., *Roseobacter denitrificans, Roseovarius* spp., *Sulfitobacter* spp., *Alpha proteobacterium* HTCC2255) and β-proteobacteria (Burkholderiales: *Bordetella* spp., *Ralstonia eutropha, Ralstonia metallidurans, Janthinobacterium* sp., and *Herminiimonas arsenicoxydans*). In Rhizobiales, the pseudoknot contains an additional nonconserved helix in a loop of the conserved pseudoknot.

In all species having several paralogs of the gene, the LEU1 pseudoknot was found upstream of no more than one paralog. The paralog was the closest homolog of *Agrobacterium tumefaciens leuA* (protein NP_355220.1) in Rhizobiales, *Roseobacter denitrificans leu*A (protein YP_681546.1) in Rhodospirillales and Rhodobacterales, and *Bordetella pertussis* Tohama I *leuA* (protein NP_879030.1) in Burkholderiales.

The LEU1 structures of Burkholderiales are similar to those of Rhizobiales and relatively distant from those of Rhodobacterales as concerns the nucleotide composition of the helices and the position of the pseudoknot relative to the stop codon of the leader peptide. The leader peptide has an absolutely nonconserved extended N-terminal region in all cases.

The *leuA* homolog that has the LEU1 regulation structure in the upstream region is not spatiall β-proteobacteria. *Magnetospirillum* spp. (Rhodospirillales) has several *leuA* paralogs. The *Magnetospirillum* paralog that is subject to the LEU1 regulation is far more distant in amino acid sequence from *E. coli leuA* than the paralog that is spatially associated with the *ilv* genes. Hence, *leuA* is probably a xenolog in Rhizobiaceae.

Thus, the LEU1 regulation is associated with the gene that belongs to the of isopropylmalate, homocitrate, and citramalate synthase family and is similar, but still distinguishable from a common *leu*A representative in most proteobacteria. On the other hand, Rhizobiaceae have only one *leuA* homolog, suggesting its functional significance, and all *leuA* homologs that have a potential LEU1 regulation in the upstream region are orthologous to each other in proteobacteria.

Magnetospirillum has at least two *leuA* paralogs coding for 2-isopropylmalate synthase. The distribution of the attenuation regulation demonstrates that one of the paralogs was inherited from a common ancestor of proteobacteria and that the other represents a LEU1-regulated xenolog. It is possible to assume that the *leuA* paralog originating from the common ancestor is regulated via the classical attenuation together with the *ilv* genes, since these genes form one operon. The situation is similar to the classical attenuation observed in the family Xanthomonadaceae of γ-proteobacteria.

The LEU1 and LEU pseudoknots are structurally similar to a certain extent, but still differ even at this level. The LEU pseudoknot has a stem, while the

LEU1 pseudoknot does not. Even a greater difference is observed at the quantitative level: the pseudoknots lack common nucleotide sequence motifs, the number of leucine codons varies from 4–8 in the LEU1 pseudoknots and is approximately 3 in the LEU pseudoknots, and the arm length of the left helix in the LEU pseudoknot is greater than in the LEU1 pseudoknot. The positions of the stop codons within a pseudoknot are similar in the two cases. In Rhodospirillales and the majority of Rhodobacterales, the pseudoknot is not far away from the start of *leuA*, and any candidate ribosome-binding site upstream of this gene is not complementary to the leucine codon region, in contrast to the LEU regulation in Actinobacteria. In Rhizobiales and Burkholderiales, the spacer between the leader peptide gene and the start of *leuA* is rather long and includes an extended nonconserved region.

If we try to search the LEU1 structure for a stem similar to the stem of the LEU structure, the attempts fail because of the following. It is possible to find a helix whose left arm overlaps the leucine codon region and has an arm length of 5–8 nt (usually 6 nt) and whose right arm overlaps the putative ribosome-binding site upstream of *leuA*, but the outer loop of such a helix will be from 140 nt in *Brucella* to 217 nt in *Bordetella*. A helix that is so long and, what is more, has an extended nonconserved region (downstream of the pseudoknot) is unstable and, consequently, seems to lack any regulatory significance. Hence, there are no grounds for believing that the LEU1 regulation functions at the level of *leuA* translation in these proteobacteria, in contrast to the LEU regulation. On the other hand, the above species additionally lack a transcription terminator in the region between the leader peptide gene and *leuA*.

Hence, a weak hypothesis is that the helices of the pseudoknot prevent the binding of a certain protein that is involved in transcription termination or degradation of the *leuA* mRNA. When the leucine concentration is low, the ribosome disrupts the pseudoknot only in rare cases, and the protein does not bind to mRNA. When the leucine concentration is high, the ribosomes translating the leader peptide unfold the pseudoknot in all cases, thus allowing the protein to bind to this region. The ribosome covers 10–12 upstream nucleotides. In the classical attenuation regulation, the distance between the regulatory codons and the antiterminator hairpin is usually comparable to the distance between the regulatory codons and the LEU1 pseudoknot in the case of the LEU1 regulation. When there is a pseudoknot, the ribosome stalling at the first leucine codons occurs at the start of the pseudoknot, but disrupts it only partly, if at all. On the other hand, when reaching the stop codon of the leader peptide, the ribosome does still not overlap the 3' arm of the right helix of the pseudoknot and does not release the proteins bound to this region.

ACKNOWLEDGMENTS

We are grateful to M.S. Gelfand for valuable discussion of the results, I. Glotova for help in computations used in this work, D. Kolobkov for the development of the Mass file, and V.V. Grechko for critical editing of the manuscript.

REFERENCES

1. 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.
2. Seliverstov A.V., Putzer H., Gelfand M.S., Lyubetsky V.A. 2005. Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiol*. **5**, 54.
3. Grundy F.J., Henkin T.M. 2003. The T box and S box transcription termination control systems. *Front Biosci*. **8**, d20–d31.
4. Grundy F.J., Henkin T.M. 2004. Regulation of gene expression by effectors that bind to RNA. *Curr. Opin. Microbiol*. **7**, 126–131.
5. Mandal M., Breaker R.R. 2004. Gene regulation by riboswitches. *Nature Rev. Mol. Cell. Biol*. **5**, 451–463.
6. Vitreschak A.G., Rodionov D.A., Mironov A.A., Gelfand M.S. 2004. Riboswitches: The oldest mechanism for the regulation of gene expression? *Trends Genet*. **20**, 44–50.
7. Das A., Crawford I.P., Yanofsky C. 1982. Regulation of tryptophan operon expression by attenuation in cell-free extracts of *Escherichia coli*. *J. Biol. Chem*. **15**, 8795–8798.
8. Henkin T.M., Yanofsky C. 2002. Regulation by transcription attenuation in bacteria: How RNA provides instructions for transcription termination/antitermination decisions. *Bioessays*. **24**, 700–707.
9. Yanofsky C. 2004. The different roles of tryptophan transfer RNA in regulating *trp* operon expression in *E. coli* versus *B. subtilis*. *Trends Genet*. **20**, 367–374.
10. Burillo S., Luque I., Fuentes I., Contreras A. 2004. Interactions between the nitrogen signal transduction protein PII and N-acetyl glutamate kinase in organisms that perform oxygenic photosynthesis. *J. Bacteriol*. **186**, 3346–3354.
11. Heery D.M., Dunican L.K. 1993. Cloning of the *trp* gene cluster from a tryptophan-hyperproducing strain of *Corynebacterium glutamicum*: Identification of a mutation in the *trp* leader sequence. *Appl. Environ. Microbiol*. **59**, 791–799.
12. Lin C., Pradkar A.S., Vining L.C. 1998. Regulation of an antranilate synthase gene in *Streptomyces venezuelae* by *trp* attenuator. *Microbiology*. **144**, 1971–1980.
13. Lyubetsky V.A., Pirogov S.A., Rubanov L.I., Seliverstov A.V. 2007. Modeling classic attenuation regulation of gene expression in bacteria. *J. Bioinform. Comput. Biol*. **5**, 155–180.
14. Chastain M., Tinoco I., Jr. 1992. Poly(rA) binds poly(rG) poly(rC) to form a triple helix. *Nucleic Acids Res*. **20**, 315–318.

15. Knorre D.G., Myzina S.D. 2000. *Biologicheskaya khimiya* (Biological Chemistry). Moscow: Nauka.

16. Semerad C.L., Maher L.J. 1994. Exclusion of RNA strands from a purine motif triple helix. *Nucleic Acids Res*. **22**, 5321–5325.

17. Carmona P., Molina M. 2002. Binding of oligonucleotides to a viral hairpin forming RNA triplexes with parallel G*GC triplets. *Nucleic Acids Res*. **30**, 1333–1337.

18. Klinck R., Guitteta E., Liquier J., Taillandier E., Gouyetteb C., Huynh-Dinhby T. 1994. Spectroscopic evidence for an intramolecular RNA triple helix. *FEBS Lett*. **355**, 297–300.

19. Holland J.A., Hoffman D.W. 1996. Structural features and stability of an RNA triple helix in solution. *Nucleic Acids Res*. **24**, 2841–2848.

20. Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Wheeler D.L. 2008. GenBank. *Nucleic Acids Res*. **36**, 25–30.

21. http://www.ncbi.nlm.nih.gov

22. Rubanov L.I., Lyubetsky V.A. 2007. RNAmodel Web Server: Modeling classic attenuation in bacteria. *In Silico Biol*. **7**, 285–308.

23. Lyubetskaya E.V., Gorbunov K.Yu. 2008. Algorithms for reconstructing evolution of regulatory signals. *Proc. 51st Sci. Conf. "Current Problems in Fundamental and Applied Sciences," Moscow Physical Technical Institute (MFTI)*. Moscow: MFTI, part 1, pp. 142–145.

24. 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.

25. Cummings L., Riley L., Black L., Souvorov A., Resenchuk S., Dondoshansky I., Tatusova T. 2002. Genomic BLAST: Custom-defined virtual databases for complete and unfinished genomes. *FEMS Microbiol. Lett*. **216**, 133–138.

26. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. **25**, 4876–4882.

27. Isambert H., Siqqia E.D. 2000. Modeling RNA folding paths with pseudoknots: Application to hepatitis delta virus ribozyme. *Proc. Natl. Acad. Sci. USA*. **97**, 6515–6520.

28. Tianbing Xia, John Santa Lucia, Jr, Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder, Xiaoqi Jiao, Christopher Cox, Douglas H. Turner. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with watson-crick base pairs. *Biochemistry*. **37**, 14719–14735.