ELSEVIER

FEMS Microbiology Letters 234 (2004) 357–370

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis

Alexey G. Vitreschak [a,*], Elena V. Lyubetskaya [a], Maxim A. Shirshin [a], Mikhail S. Gelfand [a,b], Vassily A. Lyubetsky [a,c]

[a] *Institute for Information Transmission Problems, Russian Academy of Sciences, Bolshoi Karetnyi per. 19, Moscow 127994, GSP-4, Russia*
[b] *State Scientific Centre GosNIIGenetica, 1-st Dorozhny pr. 1, Moscow 113545, Russia*
[c] *Institute for strategic stability, Atomic Energy of the Russian Federation, Luganskaya street, 9, 115304 Moscow, Russia*

Received 26 December 2003; received in revised form 16 March 2004; accepted 2 April 2004

First published online 16 April 2004

Abstract

Candidate attenuators were identified that regulate operons responsible for biosynthesis of branched amino acids, histidine, threonine, tryptophan, and phenylalanine in γ- and α-proteobacteria, and in some cases in low-GC Gram-positive bacteria, Thermotogales and Bacteroidetes/Chlorobi. This allowed us not only to describe the evolutionary dynamics of regulation by attenuation of transcription, but also to annotate a number of hypothetical genes. In particular, orthologs of *ygeA* of *Escherichia coli* were assigned the branched chain amino acid racemase function. Three new families of histidine transporters were predicted, orthologs of *yuiF* and *yvsH* of *Bacillus subtilis*, and *lysQ* of *Lactococcus lactis*. In Pasteurellales, the single bifunctional aspartate kinase/homoserine dehydrogenase gene *thrA* was predicted to be regulated not only by threonine and isoleucine, as in *E. coli*, but also by methionine. In α-proteobacteria, the single acetolactate synthase operon *ilvIH* was predicted to be regulated by branched amino acids-dependent attenuators. Histidine biosynthetic operons *his* were predicted to be regulated by histidine-dependent attenuators in *Bacillus cereus* and *Clostridium difficile*, and by histidine T-boxes in *L. lactis* and *Streptococcus mutans*.

Keywords: Regulation of gene expression; Attenuation; Branched-chain amino acids; Histidine; Threonine; Aromatic amino acids

1. Introduction

Bacteria use many different regulatory mechanisms to control transcription and translation of genes in response to concentration of metabolic products. One of possible targets for regulation is the nascent transcript during transcription elongation. Attenuation or antitermination mechanisms that involve formation of alternative RNA structures were observed in diverse bacterial groups with different molecules influencing the choice between these structures [1,2]. In enteric bacteria, many amino acid biosynthetic operons (*trp*, *his*, *leu*, *ilvGMEDA*, *ilvBN*, and *thr*) as well as the phenylalanyl-tRNA synthetase operon *pheST* are regulated by transcription attenuation [3]. This mechanism is based on coupling between transcription and translation. The nascent leader transcript contains a short open reading frame that encodes the leader peptide. Soon after transcription initiation, a secondary structure element (1:2) forms that causes RNA polymerase to pause (Fig. 1A). This pause allows the ribosome to initiate translation of the leader peptide. Then, the translating ribosome disrupts the paused complex and transcription resumes, coupled with translation. Then, two possibilities exist depending on the level of the relevant amino acid in the cell. Under the condition of amino acid starvation, the level of charged tRNA is low and it causes ribosome stalling at codons for this

* Corresponding author.
E-mail address: vitresch@loria.fr (A.G. Vitreschak).

0378-1097/$22.00 © 2004 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.femsle.2004.04.005

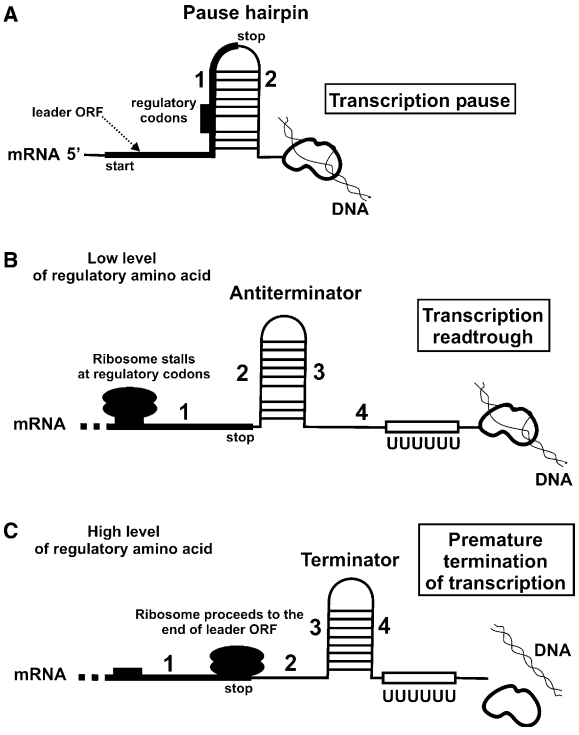

Fig. 1. The mechanism of the leader-peptide-dependent transcriptional attenuation of amino acid biosynthetic genes in bacteria. (1:2) – pause hairpin. Two alternative conformations of the 5′ UTR leader mRNA are shown, termination (1:2)/(3:4) and antitertmiantion (2:3).

amino acid (regulatory codons). When transcription proceeds, the antiterminator structure (2:3) folds and prevents terminator formation, resulting in transcription readthrough into downstream genes (Fig. 1B). Under the condition of amino acid excess, the level of charged tRNA is high and translation efficiently proceeds to the stop codon of the leader peptide. When ribosome translates the leader peptide, it prevents formation of the antiterminator structure, thereby promoting formation of the terminator (3:4), which causes premature termination of transcription (Fig. 1C). Thus, the ribosome plays the role of a mediator, sensing the concentration of charged tRNA, which in turn depends on the concentration of the amino acid. Expression of an operon corresponding to a biosynthetic pathway common for several amino acids may be regulated by all of these amino acids, and in this case the leader peptide reading frame contains several types of regulatory codons, for all amino acids.

Comparative analysis of bacterial genomes is a powerful approach to the analysis of regulation on the DNA or RNA levels and reconstruction of metabolic pathways [4–6]. Using available experimental data as a training set, we developed a program for prediction of attenuators (named LLLM [7,38]) and applied it to the analysis of upstream regions of orthologous amino acid biosynthetic genes. This resulted in identification of candidate attenuators not only in γ-proteobacteria, but in α- and β-proteobacteria, low-GC Gram-positive bacteria, as well as bacteria from some other taxa (Table 1). Analysis of regulatory peptide open reading frames allowed for prediction of the regulating amino acids. Finally, analysis of positional clustering of genes and regulatory signals leads to identification of new candidate members of the biosynthetic pathways of branched chain amino acids, histidine, threonine, and aromatic amino acids.

Three branched-chain amino acids, leucine, isoleucine and valine, are metabolically coupled in a common biosynthetic pathway, which consists of two parts (Fig. 2A). In the first part, the metabolic pathway starts from pyruvate and proceeds to valine through acetolactate synthase (IlvIH, IlvBN, and IlvGM), ketol-acid reductoisomerase (IlvC), dihydroxy-acid dehydratase (IlvD), and branched-chain amino acid aminotransferase (IlvE). Biosynthesis of leucine starts from one of the intermediates, 2-oxoisovalerate, and proceeds through 2-isopropylmalate synthase (LeuA), 3-isopropylmalate dehydratase (LeuDC), and branched-chain amino acid aminotransferase (IlvE). In the second part, the metabolic pathway starts from 2-oxobutanoate and the same proteins (IlvIH, IlvBN, IlvGM; IlvC, IlvD, and IlvE) are involved in the biosynthesis of another branched-chain amino acid, isoleucine.

In *Eschericha coli*, isoleicine, leucine, and valine biosynthetic genes (“ILV genes” below) are clustered in several operons, *ilvGMEDA*, *ilvBN*, *ilvC*, *ilvIH*, and *leuABCD* [8]. Three paralogs of acetolactate synthase are encoded by genes *ilvBN*, *ilvIH*, and *ilvGM* from three different transcriptional units. The *ilvBN* and *ilvIH* genes are transcribed as separate operons, whereas *ilvGM* is located within the *ilvGMEDA* operon. The *ilvGMEDA* and *ilvBN* operons are regulated by transcription attenuation, and the leader peptide reading frame of the attenuator contains regulatory codons for all three amino acids, isoleucine, leucine, and valine [9]. The *leuABCD* operon contains genes for the leucine biosynthesis and expression of this operon also is regulated by transcription attenuation [10]. The leader peptide of the *leu* transcription attenuator includes regulatory codons for only one amino acid, leucine. These and other operons is also regulated by repressors of transcription: *ilvC* by IlvY, *ilvIH*, and *ilvGMEDA* operons by LRP [11–14].

The histidine biosynthesis pathway consists of 10 steps and starts from 5-phosphoribosyl diphosphate, a product of the pentose phosphate pathway (Fig. 2B). The histidine biosynthesis in *E. coli* involves nine enzymes: HisGEIAFHBCD, HisF, and HisH being isozymes [15]. All genes of the histidine pathway are known to form one *his* operon regulated via transcription attenuation [16]. The leader peptide reading frame of the histidine attenuator includes a run of histidine regulatory codons.

Table 1
The list of genomes with taxonomy and abbreviations

Phylum/Class	Order	Bacteria	Abbreviations
α-Proteobacteria	Rhizobiales	*Sinorhizobium meliloti*	SM
		Agrobacterium tumefaciens	ATU
		Rhizobium leguminosarum	LE
		Mesorhizobium loti	MLO
		Bradyrhizobium japonicum	BJA
		Rhodopseudomonas palustris	RPA
		Brucella melitensis	BME
	Sphingomonadales	*Sphingomonas aromaticivorans* #	SAR
	Rhodobacteralles	*Rhodobacter sphaeroides* #	RS
	Rhodospirillales	*Magnetospirillum magnetotacticum* #	MMA
		Rhodospirillum rubrum #	RR
	Rickettsiales	*Rickettsia prowazekii*	RP
	Caulobacterales	*Caulobacter crescentus*	CO
β-Proteobacteria		*Burkholderia pseudomallei* #	BPS
		Ralstonia solanacearum	RSO
		Nitrosomonas europaea	NE
		Bordetella pertussis	BP
		Neisseria meningitidis	NM
γ-Proteobacteria	Enterobacteriales	*Escherichia coli*	EC
		Salmonella typhi	TY
		Klebsiella pneumoniae #	KP
		Erwinia carotovora	EO
		Yersinia pestis	YP
	Pasteurellales	*Haemophylus influenzae*	HI
		Pasteurella multocida	VK
		Actinobacillus actinomycetemcomitans #	AB
		Mannheimia haemolytica #	PQ
	Vibrionales	*Vibrio cholerae*	VC
		Vibrio vulnificus	VV
		Vibrio parahaemolyticus	VP
	Alteromonadales	*Shewanella oneidensis*	SH
		Microbulbifer degradans #	MDE
	Pseudomonadales	*Pseudomonas aeruginosa*	PA
		Pseudomonas putida	PP
		Pseudomonas fluorescens #	PU
		Pseudomonas syringae	PY
		Azotobacter vinelandii #	AV
		Acinetobacter spp. #	AC
	Xanthomonadales	*Xanthomonas campestris*	XCA
		Xylella fastidiosa	XFA
Firmicutes	Bacillales	*Bacillus subtilis*	BS
		Bacillus cereus	ZC
		Bacillus halodurans	HD
		Bacillus stearothermophilus #	BE
		Oceanobacillus iheyensis	OI
	Lactobacillales	*Enterococcus faecalis*	EF
		Enterococcus faecium	EFA
		Streptococcus mutans	SM
		Streptococcus pyogenes	ST
		Streptococcus pneumoniae	SPY
		Streptococcus equi #	SEQ
		Streptococcus agalactiae	SAQ
	Clostridiales	*lostridium acetobutylicum*	CA
		Clostridium perfringes	CP
		Clostridium botulinum	CB
		Clostridium difficile #	DF
		Clostridium tetani	CT
		Clostridium thermocellum	CTE
Bacteroidetes/Chlorobi		*Bacteroides fragilis*	BX
		Porphyromonas gingivalis	PFI
Thermotogae		*Thermotoga maritima*	TM
		Petrotoga miotherma	PMI
Deinococcus/Thermus		*Deinococcus radiodurans*	DR

Unfinished genomes are marked by #.

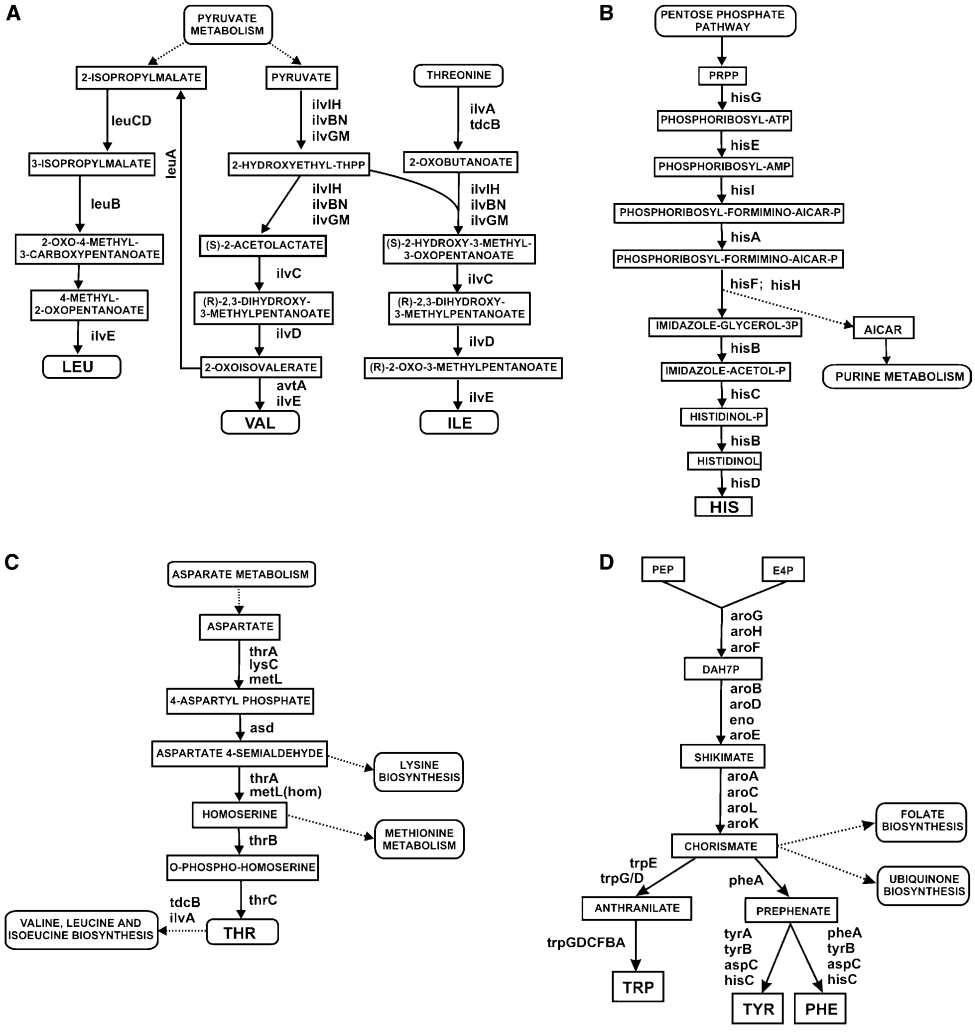

Fig. 2. Selected amino acid biosynthetic pathways of γ- and α-proteobacteria. (A) ILV (isoleucine, leucine, and valine); (B) HIS (histidine); (C) THR (threonine); and (D) aromatic amino acids (tryptophan, tyrosine, and phenylalanine).

The threonine biosynthesis is linked with biosynthesis of other amino acids, aspartate, lysine, methionine, and branched chain amino acids (Fig. 2C). A part of the pathway, which is common for threonine, methionine, and lysine biosynthesis, starts from aspartate. *E. coli* has three aspartate kinase isozymes, ThrA, MetL, and LysC, that catalyze the conversion of aspartate to 4-aspartylphosphate [17,18]. ThrA and MetL have an additional homoserine dehydrogenase (Hom) domain that catalyzes conversion of aspartate 4-semialdehyde to homoserine. The biosynthesis of branched chain amino acids starts at threonine (Fig. 2C).

In *E. coli*, expression of three isozyme genes, *thrA*, *metL*, and *lysC*, is under different regulation. Transcription of the *thrABC* operon is regulated by a threonine-isoleucine-dependent attenuator [19]. At that, regulation of the *thrABC* operon by isoleucine is an interesting example of repression by a distant product (biosynthesis of branched-chain amino acids is known to start from threonine). The aspartokinase activity of ThrA is feed-back inhibited by threonine [17]. The *metBL* operon is regulated by repressor MetJ in response of the concentration of *S*-adenosylmethionine [18]. Finally, *lysC* is possibly regulated by a lysine riboswitch LYS-element in response of the concentration of lysine (mutations in the leader region of lysC release the lysine repression in *E. coli* [20] and, moreover, LYS-element is located upstream of *lysC* [21–23]), whereas the aspartokinase activity of LysC is feed-back inhibited by lysine. Thus, the expression and activity of ThrA, MetL, and LysC isozymes are controlled by the concentration of respective amino acids.

Biosynthesis of three aromatic amino acids, tryptophan, phenylalanin, and tyrosine, is metabolically coupled (Fig. 2D) [24]. It starts with the common pathway leading from phosphoenolpyruvate and erythrose 4-phosphate through 3-deoxy-D-arabino-heptulosonate-7-phosphate and shikimate to chorismate. Then the

pathway divides into the terminal pathways, specific for each aromatic amino acid [24].

The *trp* operon of *E. coli* is regulated both by transcription attenuation and transcription repression. Transcription repressor TrpR regulates transcription initiation [25], whereas premature termination of transcription is under control of an attenuator containing two tryptophan codons [26]. The *pheA* gene, encoding chorismate/prephenate dehydratase, and *pheST* operon, encoding phenylalanyl-tRNA synthetase, are regulated by phenylanaline attenuation [27,28]. In α-proteobacterium *Rhizobium meliloti*, the *trp(E/G)* gene is known to be regulated by transcriptional attenuation [29]. In Gram-positive bacteria, tryptophan biosynthetic genes are known to be regulated by the T-box antitermination mechanism or by TRAP [30,31]. Previously we have analyzed regulation of aromatic amino acids in γ-proteobacteria [32]. Here we extend this analysis, considering newly sequenced genomes from all proteobacteria.

2. Data and methods

Complete and partial sequences of bacterial genomes were downloaded from GenBank [33]. Preliminary sequence data were obtained also from the WWW sites of The Institute for Genomic Research (http://www.tigr.org), University of Oklahoma's Advanced Center for Genome Technology (http://www.genome.ou.edu), the Sanger Centre (http://www.sanger.ac.uk), the DOE Joint Genome Institute http://www.jgi.doe.gov), and the ERGO Database [34]. The list of genomes with taxonomy and abbreviations is given in Table 1.

Protein similarity search was done using the Smith–Waterman algorithm implemented in the GenomeExplorer program [35]. Orthologous proteins were initially defined by the best bidirectional hit criterion [36] and if necessary confirmed by construction of phylogenetic trees for the corresponding protein families. The phylogenetic trees were constructed by the maximum likelihood method implemented in PHYLIP [37]. Multiple sequence alignments were done using CLUSTALW [38]. Transmembrane segments were predicted using TMpred (http://www.ch.embnet.org/software/TMPRED_form.html). The COG [36], InterPro [39] databases were used to verify the protein functional and structural annotation.

Attenuators of transcription were found using LLLM program. This program identifies candidate attenuators defined as alternative RNA hairpins such that the upstream hairpin overlaps a short open reading frame (candidate leader peptide) containing runs of regulatory codons, whereas the downstream hairpin is a candidate terminator followed by a run of Us. For details see [7,40,41].

3. Results

3.1. Isoleucine, leucine, and valine biosynthesis

Orthologs of the branched-chain amino acids (ILV) genes in genomes of γ-, β- and α-proteobacteria were identified by similarity search. Positional gene clusters corresponding to possible ILV operons are shown in Table 2. Then, the LLLM program was applied to upstream regions of the predicted ILV operons in all proteobacterial genomes. New candidate transcriptional attenuators were identified.

Attenuator-like signals were found in upstream regions of candidate *ilv* operons in γ-proteobacteria (Enterobacteria, Pasteurellales, Vibrionales, *Shewanella oneidensis*, and Xanthomonadales). In Pseudomonadales and other bacteria, the *ilv* genes are scattered along a genome, and some of them are also preceded by candidate attenuators. The *ilvBN* operon, which encodes genes for one of the acetolactate synthase isozymes in Enterobacteria, also was predicted to be regulated by the attenuation mechanism via leucine and valine regulatory codons. Other predicted attenuators include regulatory codons for three amino acids, isoleucine, leucine, and valine, similar to the experimentally studied attenuators of *E. coli* (Fig. 3).

The structure of the candidate *ilv* biosynthetic operons varies in the analyzed genomes. For example, the order of genes in the *ilv* operon is *ilvGMEDA* in Enterobacteria and Vibrionales, but in Xanthomonadales, the order is *ilvCGM–tdcB–leuA*. In the latter case, the *tdcB* gene is possibly co-regulated with the ILV genes. Its product is threonine dehydratase which catalyzes reactions in both serine and ILV metabolism.

Another possible co-regulation event was observed in *Pasteurella multocida*. A gene with unknown function (orthologous to hypothetical gene *ygeA* of *E. coli*) is located within the *ilv* operon (*ilvGM–ygeA–DA*), and a candidate attenuator was found upstream of this operon. YgeA is weakly similar to the amino acid racemase protein RacX from *B. subtilis*, which converts L-aspartate to D-aspartate [42,43]. Thus, *ygeA* likely encodes a new kind of racemase, possibly ILV racemase.

The *leu* operon, which includes only genes for the leucine synthesis, is predicted to be regulated by attenuation in some γ-proteobacteria (Enterobacteria, Pasteurellales, Vibrionales, Alteromonadales, and Xanthomonadales), but not in Pseudomonadales and other species. The leader peptide reading frames of all predicted attenuators include runs of leucine codons.

Little is known about regulation of ILV genes in α-proteobacteria. Expression of the *ilvIH* genes encoding the two subunits of acetolactate synthase has been studied in *Caulobacter crescentus*, and the region between *ilvIH* and the transcription initiation site was

Table 2
Predicted operon structures and regulation of the ILV genes

α	*Sinorhizobium meliloti*	*leuB; leuC; leuD; %leuA2; ilvE; ilvD; ilvA; **REG2**-ilvC; &ilvIH;*
	Agrobacterium tumefaciens	*leuDB; leuC; %leuA2; ilvE; ilvD; ilvA; **REG2**-ilvC &ilvIH;*
	Mesorhizobium loti	*leuD-//-leuB; leuC; %leuA2; ilvE; ilvD; ilvA; **REG2**-ilvC &ilvIH;*
	Bradyrhizobium japonicum	*leuB; leuC; leuD; %leuA; ilvE; ilvD; ilvA; ilvC; &ilvI-x-H;*
	Rhodopseudomonas palustris	*leuB; leuC-//-leuD; %leuA; leuA2; ilvE; ilvD; ilvA; &ilvIH-//-ilvC;*
	Brucella melitensis	*leuB; leuC; leuD; %leuA2; ilvE; ilvD; ilvA; &ilvIH-//-ilvC;*
	Sphingomonas aromaticivorans #	*leuC-//-D; -leuB-; %leuA2; ilvE; ilvD; ilvA; ilvIHC;*
	Rhodobacter sphaeroides #	*leuCD; leuB; %leuA; ilvE; ilvD; ilvA; ilvC; &ilvIH;*
	Magnetospirillum magnetotacticum #	*leuCDB; %leuA; leuA2; ilvE; ilvD; ilvA; &ilvIHC-//-leuA;*
	Rhodospirillum rubrum #	*leuCDB; leuA2; leuA2; ilvE; ilvD; ilvA; ilvIHC;*
	Rickettsia prowazekii	-
	Caulobacter crescentus	*leuCD-x-B; %leuA; ilvE; ilvD; ilvA; ilvC; &ilvIH;*
β	*Burkholderia pseudomallei #*	*leuCDB leuA2; ilvE; ilvD; ilvA; ilvIHC-//-%leuA;*
	Ralstonia solanacearum	*leuCDB; leuA2; ilvE; ilvD; ilvA; ilvIHC-//-%leuA;*
	Nitrosomonas europaea	*leuCDB; ilvE; ilvD; ilvA; ilvIHC-//-leuA;*
	Bordetella pertussis	*leuCDB; leuA2; %leuA2; ilvE; ilvD; ilvA; ilvIHC;*
	Neisseria meningitidis	*leuCD; leuB; leuA; ilvE; ilvD; ilvA; ilvIHC;*
γ	*Escherichia coli*	***leuO***<-->*&leuABCD; &ilvGMEDA **ilvY**<->ilvC; &ilvBN; ilvIH;*
	Salmonella typhi	***leuO***<-->*&leuABCD; &ilvGMEDA **ilvY**<->ilvC; &ilvBN; ilvIH;*
	Klebsiella pneumoniae #	***leuO***<-->*&leuABC[&ilvGMEDA **ilvY**<->ilvC; &ilvBN;*
	Yersinia pestis	***leuO***<-->*&IS-leuABCD; &ilvGMEDA **ilvY**<->ilvC; &ilvBN; ilvIH;*
	Erwinia carotovora	*&leuABCD; &ilvGMEDA **ilvY**<->ilvC; &ilvBN; ilvIH;*
	Haemophylus influenzae	*&leuABCD; ilvGDA; ilvE; **ilvY**<->ilvC; ilvIH;*
	Pasteurella multocida	*&leuABCD; &ilvGM-ygeA-DA; ilvE; **ilvY**<->ilvC; ilvIH;*
	Mannheimia haemolytica #	*&leuA; leuB;]leuCD; ilvGM; -ilvE-; [ilvD[; ilvA-; ilvC; ilvIH;*
	Vibrio cholerae	***leuO***<-->*&leuABCD; &ilvGMEDA; **ilvY**<->ilvC; ilvIH;*
	Vibrio vulnificus	***leuO***<-->*&leuABCD; &ilvGMEDA; **ilvY**<->ilvC; ilvIH;*
	Vibrio parahaemolyticus	***leuO***<-->*&leuABCD; &ilvGMEDA; **ilvY**<->ilvC; ilvIH;*
	Shewanella oneidensis	*&leuABCD; &ilvGMDA; -ilvE-; **ilvY**<->ilvC; ilvIH;*
	Pseudomonas aeruginosa	***REG***<-->*leuCD-yafE-leuB; leuA2; -ilvE-; ilvD; ilvD; X-ilvA; ilvA; ilvIHC;*
	Pseudomonas putida	***REG***<-->*leuCD-yafE-leuB; leuA2; ilvE; ilvD; &ilvA; ilvA; ilvIHC;*
	Pseudomonas fluorescens	***REG***<-->*leuCD-x-B; leuA2; ilvE; ilvD; &ilvA; ilvA; ilvIHC;*
	Pseudomonas syringae	***REG***<-->*leuCDB; leuA2; ilvE; ilvD; &ilvA; ilvIHC;*
	Acinetobacter spp.	***REG***<-->*leuCxDB; leuA2; -ilvE-; ilvD; ilvD; ilvA; ilvIHC;*
	Azotobacter vinelandii #	***REG***<-->*leuCDB; -leuA- -ilvE-; ilvD; ilvD; ilvA; ilvIHC;*
	Microbulbifer degradans #	***REG***<-->*leuCDB; -leuA- -ilvE-;]ilvD **ilvY**<->ilvC; ilvA; ilvIH;*
	Xanthomonas campestris	***REG***<-->*leuCD-yafE-leuB &ilvCGM-tdcB-leuA; ilvA; ilvE; ilvD;*
	Xylella fastidiosa	*leuCDB; &ilvCGM-tdcB-leuA; ilvE; ilvD;*
D/T	*Deinococcus radiodurans*	*leuC; leuDB; &leuA; -ilvE-; ilvD; -ilvA; &ilvIHC;*

Predicted attenuators are denoted by '&' and '%' (the latter lack terminator-like RNA secondary structures, see the text). Divergently located genes are separated by '<->'. Contig ends are marked by square brackets. Known and possible (REG) regulators from LysR family are shown in bold. Genes with unknown function are denoted by 'x' (with numbers for orthologous genes).

shown to have the properties of a transcription attenuator [44] (in the cited paper this operon is called *ilvBN*, not *ilvIH*, but phylogenetic analysis of all three acetolactate synthases shows that this gene is located on the branch corresponding to *ilvIH*, data not shown). We analyzed upstream regions of all ILV genes of available α-proteobacterial genomes and found attenuator-like structures (Table 2). α-Proteobacteria have one acetolactate synthase, IlvIH. The *ilvIH* operon is possibly regulated by transcription attenuation in Rhizobiales (*Sinorhizobium meliloti*, *Agrobacterium tumefaciens*, *Mesorhizobium loti*, *Bradyrhizobium japonicum*, *Rhodopseudomonas palustris*, and *Brucella melitensis*), *Rhodobacter* spp., *Magnetospirillum magnetotacticum*, *C. crescentus*, and a deeply rooted bacterium *Deinococcus radiodurans* (Deinococcus/Thermus group). The leader peptide reading frames of predicted attenuators include runs of isoleucine, leucine, and valine regulatory codons (Fig. 3). Conversely, in γ-proteobacteria, operons encoding two other acetolactate synthase isoenzymes, *ilvBN* (present only in Enterobacteria) and *ilvGM*, but not *ilvIH*, are regulated by attenuators.

There exist two groups of homologous 2-isopropylmalate synthases, *leuA* and *leuA2* (approx. 30% sequence identity). The *leuA* genes, orthologs of *leuA* from *E. coli* were observed in γ-proteobacteria, excluding Pseudomonadales, and in some α-proteobacteria, whereas the *leuA2* genes, homologs of well-studied 2-isopropylmalate synthases from Actinobacteria and Fungi, in particular *Corynebacterium glutamicum* [45] and *Saccharomyces cerevisiae* [46], respectively, were observed in α-proteobacteria, some β-proteobacteria and Pseudomonadales. In α-proteobacteria, both types of 2-isopropylmalate synthase genes have candidate attenuators in upstream regions (Table 2). Although these attenuators have leader peptide reading frames with runs of leucine regulatory codons, the terminator structures are weak and lack runs of uridines (Fig. 3). At that, one should note that a similar situation was observed in the case of *trpE* and *trpGDC* operons in *Pseudomonas putida*, where transcripts were attenuated despite the absence of strong ρ-independent terminator structures [47]. Moreover, we found a possible attenuator with a strong G/C-rich terminator upstream of the *leuA* gene in *D. radiodurans*.

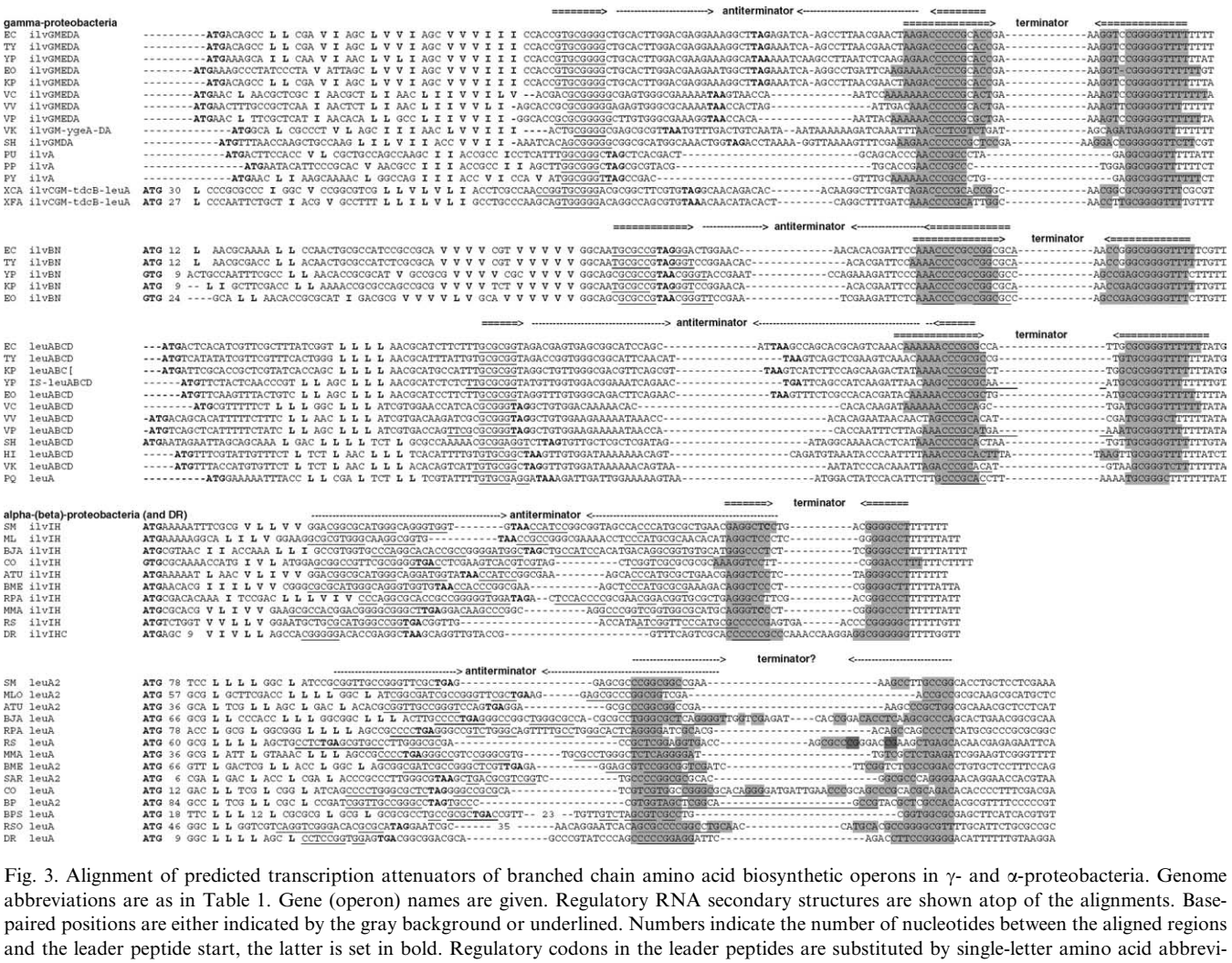

Fig. 3. Alignment of predicted transcription attenuators of branched chain amino acid biosynthetic operons in γ- and α-proteobacteria. Genome abbreviations are as in Table 1. Gene (operon) names are given. Regulatory RNA secondary structures are shown atop of the alignments. Base-paired positions are either indicated by the gray background or underlined. Numbers indicate the number of nucleotides between the aligned regions and the leader peptide start, the latter is set in bold. Regulatory codons in the leader peptides are substituted by single-letter amino acid abbreviations: I (isoleucine), L (leucine), and V (valine).

3.2. Histidine biosynthesis

Orthologs of the histidine biosynthetic (HIS) genes in bacterial genomes were identified by similarity search. Positional gene clusters corresponding to candidate HIS operons are listed in Table 3. The LLLM program with parameters obtained by analysis of known attenuator structures was used to scan the upstream regions of predicted HIS operons in all analyzed genomes (for details see [7]). New candidate transcriptional attenuators were identified, mainly in γ-proteobacteria. We also identified attenuator-like structures in some low-GC Gram-positive bacteria, Bacteroidetes/Chlorobi group and *Thermotogales*.

Positional analysis and analysis of regulation showed that in most γ-proteobacteria (Enterobacteria, Pasteurellales, Vibrionales, and *Shewanella oneidensis*), all histidine biosynthetic genes are clustered and possibly regulated via the transcription attenuation mechanism (Table 3). All candidate attenuators upstream of the *his* operons in these bacteria have similar features: a leader peptide reading frame with a run of histidine regulatory codons and terminator/antiterminator structures (Fig. 4). We found no attenuators upstream of HIS genes in Pseudomonadales, Xanthomonadales, and some other γ-proteobacteria.

Analysis of upstream regions of HIS genes in other taxonomic groups revealed attenuator-like structures in the *Bacillus*/*Clostridium* group, Bacteroidetes/Chlorobi, and Thermotogales. In those cases, histidine biosynthetic operons, which include most of HIS genes, are possibly regulated. We observed diversity of mechanisms for regulation of the HIS gene expression. In particular, in *Lactococcus lactis* and *Streptococcus mutans*, the *his* operon is regulated by the T-box antitermination mechanism [[48], Vitreschak A, unpublished], whereas in *Bacillus cereus* and *Clostridium difficile*, the *his* operon seems to be regulated via transcription attenuation. Other *Streptococcus* spp. as well as *Enterococcus* spp. lack histidine biosynthetic genes. Moreover,

Table 3
Candidate operon structures and predicted regulation of HIS genes. Notation as in Table 2

γ	*Escherichia coli*	*&hisGDCBHAF(I/E);*
	Salmonella typhi	*&hisGDCBHAF(I/E);*
	Klebsiella pneumoniae #	*&hisGDCBHAF(I/E);*
	Yersinia pestis	*&hisGDCBAHAF(I/E);*
	Erwinia carotovora	*&hisGDCBHAF(I/E);*
	Vibrio cholerae	*&hisGDCBHAF(I/E);*
	Vibrio vulnificus	*&hisGDCBHAF(I/E);*
	Vibrio parahaemolyticus	*&hisGDCBHAF(I/E);*
	Shewanella oneidensis	*&hisGDCBHAF(I/E);*
	Haemophylus influenzae	*&hisGDCBHAF(I/E); &HI0325;*
	Pasteurella multocida	*&hisG-x-vatB-DCBHAF(I/E);*
	Actinobacillus actinomycetemcomitans #	*his(I/E);*
	Mannheimia haemolytica #	*]hisGD; &x-actX2-hisCB[ ]hisH-x-AF-x-(I/E);*
	Pseudomonas aeruginosa	*-hisGDC; his(B1)H-x1-AF; hisIE;*
	Pseudomonas putida	*-hisGDC; his(B1)H-x1-AF; hisIE;*
	Pseudomonas fluorescens #	*hisGDC; his(B1)H-x1-AF; hisIE;*
	Pseudomonas syringae	*hisGDC; his(B1)H-x1-AF; hisIE;*
	Azotobacter vinelandii #	*-hisGDC; his(B1)H-x1-AF; hisIE;*
	Acinetobacter spp. #	*-hisG[ ]hisDC; his(B1)H-x-A; hisF; his(I/E);*
	Microbulbifer degradans #	*-hisGD; hisC; his(B1)HAF; hisIE;*
	Acidithiobacillus ferrooxidans #	*-hisGDC(B1)HAF(I/E);*
	Xanthomonas campestris	*x2-hisGDCBHAF(I/E);*
	Xylella fastidiosa	*x2-hisGDCBHAF(I/E);*
α	*Caulobacter crescentus*	*&hisZ-x-hisG; x-his(B1)HAFE; -hisD-; hisC; hisI-;*
B/C	*Bacillus cereus*	*&hisZ-hisGD(B1)HAFIE-hisB2; x-hiB2; hisC; -hisC-; &hisZ2; ThisS-aspS; &yvsH1; LyvsH2;*
	other Bacillus sp. (BS, BE, HD)	*hisZ-hisGD(B1)HAF(I/E); -hisC-; ThisS-aspS;*
	Clostridium difficile #	*&hisZ-hisGC(B1)HAF(I/E); hisD; hisB2-x; hisB2; hisB2; -hisS-aspS;*
	Lactococcus lactis	*ThisC-hisZ-hisGD-x-(B1)-x-HAF(I/E); ThisS-aspS; &lysQ; LlysP;*
	Streptococcus mutans	*ThisC-hisZ-hisGD-x-(B1)-x-HAF(I/E); hisS-aspS;*
	other Streptococcus sp. (SP, SPY, SEQ, SAG) Enterococcus sp. (EF, EFA)	*- hisS-aspS;*
Therm	*Thermotoga maritima*	*&IS-hisS-hisGDC(B1)HAF(I/E); -hisB2;*
	Petrotoga miotherma #	*&hisS-hisGDC(B1)HAF(I/E); -hisB2;*
CFB	*Polaribacter filamentus #*	*&hisGDCBHAF-actX3-(I/E)*
	Bacteroides fragilis	*&hisGDCB; hisHAF(I/E);*

Gene fusion of *hisI* and *hisE* is denoted by (I/E). Histidine-specific T-boxes are denoted by 'T'. LYS-elements are denoted by 'L'.

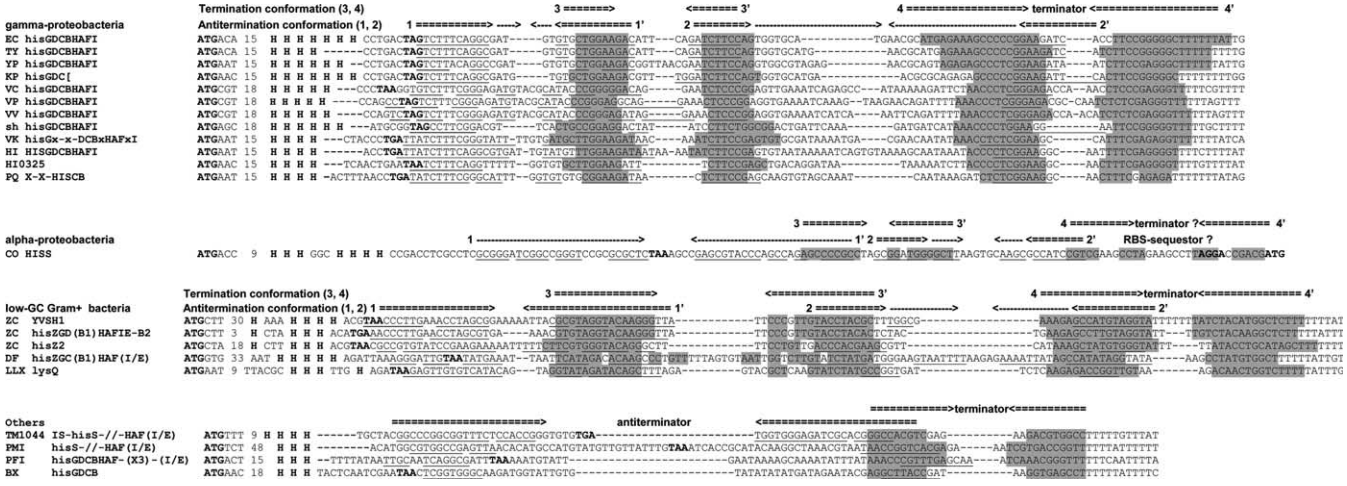

Fig. 4. Alignment of predicted transcription attenuators of histidine biosynthetic operons in various bacteria. Notation as in Fig. 3. H denotes histidine regulatory codons in the leader peptide reading frame.

B. cereus has two copies of the *hisZ* gene, which are predicted to be regulated by transcriptional attenuation: one as a part of the *his* biosynthetic operon; the other, as a separate gene with a possible histidine attenuator structure in the upstream region (Table 3). *hisS* gene in this bacterium, as well as orthologous *hisS* genes in *Bacillus* spp., *Listeria* spp., *Enterococcus* spp., and *L. lactis*, are located separately and predicted to be regulated by the T-box antitermination mechanism [[49], Vitreschak A, unpublished].

Several hypothetical genes were predicted to belong to the histidine regulons. *HI0325* from *Haemophylus influenzae*, which encodes a putative transporter with 10 transmembrane segments, has a candidate histidine attenuator in the upstream region. This gene is widely distributed, but not universal in bacteria. In a number of genomes, in particular in *Fusobacterium nucleatum* and *Bacillus halodurans*, this gene is clustered with histidine utilization genes (the *hut* locus). Thus, *HI0325* and its orthologs (*yuiF* in *B. subtilis*) possibly constitute a new family of histidine transporters.

Another example is the *BC0629* gene from *B. cereus* that also is possibly regulated via the histidine-dependent attenuation. This gene (*yvsH* in *B. subtilis*) is homologous to the arginine:ornithine antiporter *arcD* from *Pseudomonas aeruginosa* and lysine permease *lysI* from *Corinobacterium glutamicum.* All these proteins belong to the basic amino acid/polyamine antiporter APA family [http://tcdb.ucsd.edu/tcdb/background.php]. *B. cereus* has two *yvsH* paralogs, *yvsH1* (*BC0629*) and *yvsH2* (*BC0865*). The former is a candidate lysine transporter whose expression was predicted to be regulated by the lysine via the *LYS*-element riboswitch mechanism [21]. The upstream region of *yvsH2* contains a candidate attenuator whose leader peptide reading frame contains a run of histidine regulatory codons (Fig. 4). Thus, *yvsH2* (*BC0629*) is possibly involved in the histidine transport. The predicted specificity of this transporter is consistent with experimental data for the homologous HisJ and LAO transporters, which both bind histidine, arginine, lysine, and ornithine, albeit with different affinities towards these ligands [50].

A very similar situation was observed in the case of two paralogous transporter genes in *L. lactis*, *lysP*, and *lysQ.* Both proteins are similar (more than 50% identity) to the experimentally identified lysine permease *lysP* of *E. coli* [51]. In the *L. lactis* genome, *lysP* was predicted to be regulated by a *LYS*-element and thus to be involved in the lysine transport [21]. On the other hand, the upstream region of *lysQ* contains a candidate histidine attenuator (Fig. 4). Thus, these two transporters can have different affinity to lysine and histidine, and because of that be regulated one by lysine and the other one by histidine.

All genes required for the histidine biosynthesis were identified in all analyzed bacteria, the only exception being the histidinol-phosphatase domain of HisB in *Pseudomonas* spp. Neither similarity search nor positional analysis and analysis of regulation provided a candidate for this enzymatic activity.

On the other hand, at least three non-homologous proteins with unknown function (shown in Table 3 as vatB, actX2, and actX3 in *P. multocida*, *Mannheimia haemolytica*, and *Polaribacter filamentus*, respectively), encoding putative acetyltransferases, that are possibly co-regulated with HIS genes. These candidate acetyltransferases could catalize conversion of histamine to 4-β-acetylaminoethyl-imidazole. This is one of the steps of the histidine modification (http://www.genome.ad.jp/kegg/metabolism.html), for which only enzymatic activity, EC 2.3.1., is known, but no genes have been assigned yet.

3.3. Threonine biosynthesis

We analyzed regulation of the *thr* biosynthetic operon in proteobacteria. Orthologs of *thr* genes were identified by similarity search. Candidate *thr* operons and possible regulation are shown in Table 4. Enterobacteria, Pasteurellales, Vibrionales, *Shewanella oneidensis*, and Xanthomonadales have the same gene order *thrABC* in the threonine biosynthetic loci. In Pseudomonadales and some other genomes, the threonine biosynthetic genes are scattered along genome. Moreover, in Enterobacteria, Pasteurellales, Vibrionales, *S. oneidensis*, and Xanthomonadaels, *thrA* encodes a bifunctional protein, aspartate kinase/homoserine dehydrogenase, whereas in Pseudomonadales and some other γ-proteobacteria *thrA2* (aspartate kinase) and *hom* (homoserine dehydrogenase) are located in different loci. Finally, two homoserine kinase genes, *thrB2* and *thrH* [52], neither homologous to *thrB* of *E. coli*, were observed in Pseudomonadales (Table 4).

Then, we analyzed upstream regions of the predicted *thr* operons by LLLM trained on known attenuators. New candidate transcriptional attenuators were identified in γ-proteobacteria (Table 4). They have all properties of threonine attenuators: a short leader peptide reading frame with a run of threonine and isoleucine codons, as well as alternative termination and antitermination RNA structures (Fig. 5). Our results predicted that *thr* operons are regulated by transcription attenuation in Enterobacteria, Pasteurellales, Vibrionales, *Shewanella oneidensis*, and *Xanthomonas campestris.*

Closer analysis showed that in Pasteurellales (*H. influenzae*, *P. multocida*, *Actinobacillus actinomycetemcomitans*, and *M. haemolytica*), the leader peptide reading frame contains not only standard regulatory codons for threonine and isoleucine, but also numerous methionine codons (Fig. 5). Thus, the *thr* operons in Pasteurellales seem to be regulated by concentration of three amino acids, threonine, isoleucine, and methionine, instead of the former two. Indeed, Pasteurellales have only one copy of the bifunctional aspartate kinase/homoserine dehydrogenase protein, instead of two isozymes ThrA and MetL in other γ-proteobacteria, where the expression of these isozymes is regulated by threonine/isoleucine and by methionine, respectively. Thus, it makes sense that the single ThrA isozyme of Pasteurellales is regulated not only by threonine and isoleucine, but by methionine as well. One more, monofunctional aspartate kinase LysC, is present in three of the five

Table 4
Candidate operon structures and predicted regulation of THR genes

γ	*Escherichia coli*	*&thrABC;*
	Salmonella typhi	*&thrABC;*
	Klebsiella pneumoniae #	*&thrABC;*
	Yersinia pestis	*&thrABC;*
	Erwinia carotovora	*&thrABC;*
	Vibrio cholerae	*&thrABC; thrA2;*
	Vibrio vulnificus	*&thrABC;*
	Vibrio parahaemolyticus	*&thrABC;*
	Shewanella oneidensis	*&thrABC; thrA2;*
	Haemophylus ducrei	-
	Haemophylus influenzae	*&thrABC;*
	Pasteurella multocida	*&thrABC;*
	Actinobacillus actinomycetemcomitans #	*&thrAB-x-C;*
	Mannheimia haemolytica #	*&thrAB;]thrC;*
	Pseudomonas aeruginosa	*thrA2; hom-thrC; -thrB2; thrH;*
	Pseudomonas putida	*thrA2; hom-thrC; -thrB2;*
	Pseudomonas fluorescens #	*thrA2; thrA2; hom-thrC; -thrB2; thrH;*
	Pseudomonas syringae	*thrA2; hom-thrC; -thrB2; thrH;*
	Acinetobacter spp. #	*thrA2; hom] thrB2; thrH;*
	Azotobacter vinelandii #	*thrA2; -hom-; -thrC-; x-thrB2;*
	Microbulbifer degradans #	*thrA2; thrA2-; hom-thrC; thrH;*
	Xanthomonas campestris	*&thrAB-thrC;*
	Xylella fastidiosa	*thrABC;*

Notation as in Table 2.

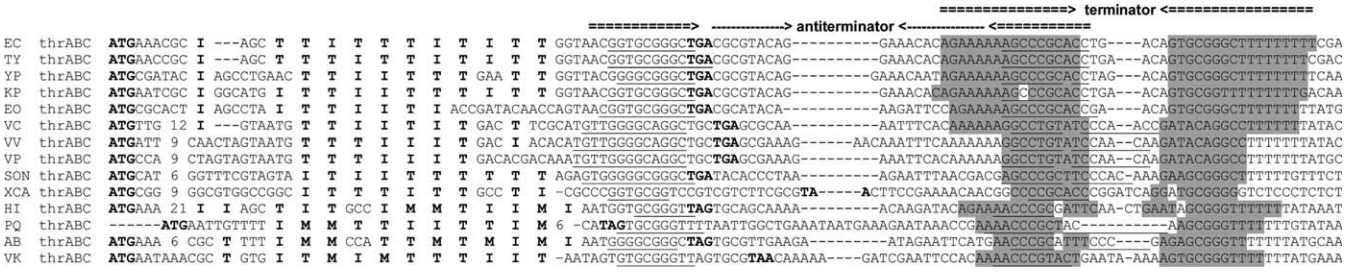

Fig. 5. Alignment of predicted transcription attenuators of *thr* operons in γ-proteobacteria. Notation as in Fig. 3. T, I, and M denote, respectively, threonine, isoleucine, and metionine regulatory codons in the leader peptide reading frame.

Pasteurellales, *P. multocida*, *Haemophylus ducrei*, and *M. haemolytica*, and the expression of *lysC* has been predicted to be regulated by lysine via *LYS*-element riboswitches, as in *E. coli* [21–23].

3.4. Tryptophan and phenylalanine biosynthesis

Orthologs of the *trp* and *pheA* genes in γ- and α-proteobacteria were identified by similarity search. Candidate *trp*, *pheA*, and *pheST* operons are shown in Table 5. Candidate attenuators were identified upstream of these operons by the LLLM program (Table 5).

Candidate *trp* attenuators found in Enterobacteria, Vibrionales, and *Shewanella oneidensis* have leader peptide reading frames with tryptophan regulatory codons and antitermination/termination-like structures (Fig. 6). The *trp(E/G)* gene, which encodes fused components of anthranilate synthase responsible for the first step of the tryptophan biosynthesis, is possibly regulated by transcription attenuation in all analyzed Rhizobiales (order of α-proteobacteria) and in *Bordetella pertussis* (belonging to β-proteobacteria).

The *pheA* operon may be regulated by candidate phenylalanine-dependent attenuators in Enterobacteria, Vibrionales, and *S. oneidensis*, whereas the *pheST* operon seems to be regulated only in Enterobacteria.

Candidate attenuators of the *trpE* and *trpGDC* operons in Pseudomonadales have some peculiar properties. There is experimental evidence that transcription of the *trpE* and *trpGDC* operons is regulated by attenuation [47], but no strong ρ-independent transcriptional terminators could be found in the leader regions of these operons. We aligned sequences upstream of the *trpE* and *trpGDC* operons from five pseudomonads. The region

Table 5
Candidate operon structures and predicted regulation of *trp* and *pheA* genes

α Rhizobiales	*Mesorhizobium loti*	&trp(E/G);	-trpDC-	trpFBA;	-pheA;	pheS-x-T;
	Brucella melitensis	&trp(E/G);	-trpDC-	trpFBA;	-pheA;	pheST;
	Sinorhizobium meliloti	&trp(E/G);	-trpDC-	trpFBA;	-pheA;	pheST;
	Agrobacterium tumefaciens	&trp(E/G);	-trpDC-	trpFBA;	-pheA;	pheST;
	Rhodopseudomonas palustris #	&trp(E/G);	-trpDC-	trpFBA;	-pheA;	pheS-x-T;
	Rhizobium leguminosarum #	&trp(E/G);	-trpDC-	trpFBA;	]pheA;	pheS[;
	Bradyrhizobium japonicum	&trp(E/G);	-trpDC	trpFBA;	-pheA;	pheST;
β	*Bordetella pertussis*	&trpEGDC;			pheA;	pheST;
γ	*Escherichia coli*	&trpE(G/D)(C/F)BA;			&pheA;	&pheST;
	Salmonella typhi	&trpE(G/D)(C/F)BA;			&pheA;	&pheST;
	Klebsiella pneumoniae #	&trpEGD(C/F)BA;			&pheA;	&pheST;
	Yersinia pestis	&trpEGD(C/F)BA;			IS-pheA;	&pheST;
	Erwinia carotovora	&trpEGD(C/F)BA;			&pheA;	&pheST;
	Haemophylus influenzae	trpEG-x1-trpD(C/F)-		-trpBA;	pheA;	pheST;
	Pasteurella multocida	trpEG<-x-D(C/F)-x-trpBA;		trpBA;	pheA;	pheS-x-T;
	Actinobacillus actinomycetemcomitans #	trpEG-x1-trpD(C/F)-		-trpBA;	pheA;	pheS-x-T;
	Mannheimia haemolytica #	trpEG-x1-trpD)-	trp(C/F)[	-trpBA;	pheA;	pheST;
	Vibrio cholerae	&trpEGD(C/F)BA;			&pheA;	pheST;
	Vibrio vulnificus	&trpEGD(C/F)BA;			&pheA;	pheST;
	Vibrio parahaemolyticus	&trpEGD(C/F)BA;			&pheA;	pheST;
	Shewanella oneidensis	&trpEGD(C/F)BA;			&pheA;	pheST;
	Pseudomonas aeruginosa	%trpE;	%trpGDC;	-trpF; trpI<->trpBA;	-pheA-	pheST;
	Pseudomonas putida	%trpE;	%trpGDC;	-trpF; trpI<->trpBA;	-pheA-	pheST;
	Pseudomonas fluorescens #	%trpE;	%trpGDC;	-trpF; trpI<->trpBA;	-pheA-	pheST;
	Pseudomonas syringae	%trpE;	%trpGDC;	-trpF; trpI<->trpBA;	-pheA-	pheST;
	Azotobacter vinelandii #	%trpE;	%trpGDC;	-trpF; trpI<->trpBA;	-pheA-	pheST;
	Acinetobacter spp. #	trpE;	trpGDC;	trpFBA;	pheA-	pheST;
	Microbulbifer degradans #	trpE;	trpGDC;	trpFBA-	-pheA-	pheST;
	Acidithiobacillus ferrooxidans #	trpEGDC-;		-trpF; trpBA;	]pheA-	]pheST[;
	Xanthomonas campestris	trpE;	trpGxDC-	-trpF; trpI<->trpB-x-trpA;	-pheA-	pheST;
	Xylella fastidiosa	trpEGDC-x2;		-trpFBA	-pheA-	pheST;

Notation as in Table 2. Gene fusion of *trpE* and *trpG* is denoted by *(E/G)*. Gene fusion of *trpC* and *trpF* is denoted by *(C/F)*.

of sequence conservation corresponds to a possible leader peptide, which contains two nearly adjacent tryptophan codons (Fig. 6). It seems that in this case the terminator and antiterminator structures are less pronounced and maybe less stable than those in other attenuators.

4. Discussion

This analysis allowed us to identify a large number of candidate attenuators and predict the amino acid(s) responsible for the regulation, demonstrated variability of regulatory mechanisms for the amino acid biosynthetic pathways even in closely related genomes, and allowed for functional annotation of hypothetical genes encoding transporters and enzymes. In particular, candidate attenuators were found in some taxonomic groups where this mechanism of regulation was studied little (α-proteobacteria, low-GC Gram-positive bacteria) or not at all (Bacteroidetes/Chlorobi group and Thermotogales).

This analysis, as well as other comparative studies, demonstrate the diversity and evolutionary lability of regulatory mechanisms based on formation of alternative RNA structures, especially in low-GC Gram-positive bacteria. Indeed, we observed candidate histidine attenuators regulating *his* operons in bacilli and clostridiae, but T-boxes in streptococci that have this operon. It is known that transcription attenuation and T-box antitermiantion mechanisms are prevalent in Proteobacteria and Gram-positive bacteria, respectively. We demonstrate that these different mechanisms, based on switching between two conformations of the RNA nascent transcript, are involved in regulation of the *his* operons in low-GC Gram-positive bacteria. For example, candidate histidine attenuators regulate *his* operons in *B. cereus* and *C. difficile*, but not in *L. lactis* and *S. mutans*, where this role is taken by histidine T-boxes. Moreover, in *B. cereus* both regulatory mechanisms are present, where histidine attenuators regulate two operons *his* and *hisZ2*, whereas the third one, *hisS*, is regulated by a histidine T-box. This situation is similar to the one with the methionine biosynthesis pathway, which is regulated by T-boxes in streptococci, S-box riboswitches in bacilli and clostridiae, and by transcription repression in lactobacilli [53].

In the case of transcription attenuation, we suppose an ancient origin of this regulatory mechanism. Indeed, we found possible attenuators of amino acid biosynthetic genes not only in proteobacteria, but also in low-GC Gram-positive bacteria, *Bacteroidetes/*

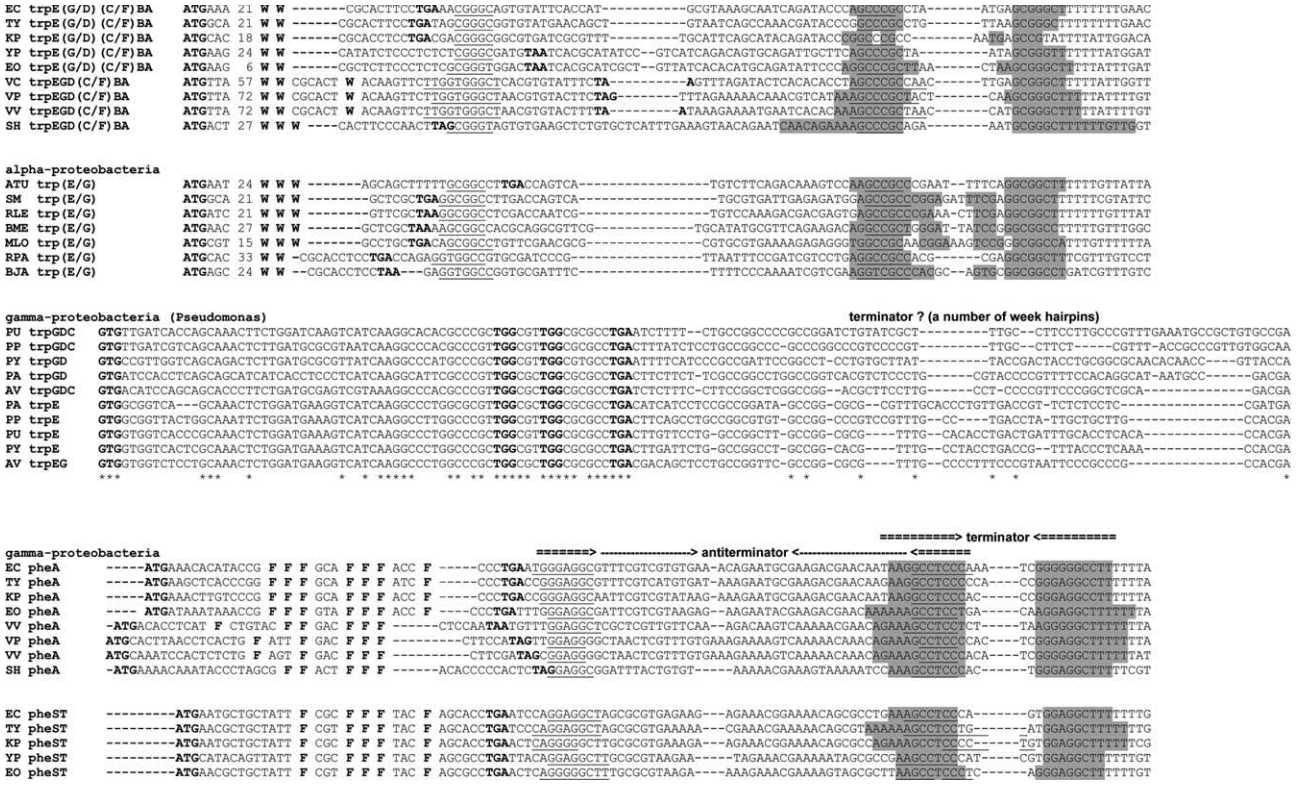

Fig. 6. Alignment of predicted transcription attenuators of *trp*, *pheA*, and *pheST* operons in γ- and α-proteobacteria. Notation as in Fig. 3. W and F denote tryptophan and phenylalanine regulatory codons, respectively, except in *trp* operons of *Pseudomonas*, where tryptophan codons TGG are retained.

Chlorobi, and, notably, in deeply rooted bacteria, Thermotogales and *D. radiodurans*. The hypothesis of the ancient origin of transcription attenuation and some others regulatory mechanisms, based on formation of alternative RNA structures, is reasonable. In fact, a number of riboswith elements involved in regulation of genes from various metabolic pathways (vitamin, purine, lysine, and methionine biosynthesys and transport) were identified in a large number of distant bacteria [for review see [53].

Candidate threonine/isoleucine-dependent attenuators were found upstream of *thr* operons in Enterobacteria, Pasteurellales, Vibrionales, *Shewanella oneidensis*, and Xanthomonadales. In Pasteurellales, attenuators of the *thr* operon were predicted to respond not only to the level of threonine and isoleucine, but also to methionine. Thus, the single bifunctional aspartate kinase/homoserine dehydrogenase ThrA of these species is regulated by all three amino acids. In fact, probable regulation of *thrA* in Pasteurellales by not only threonine and isoleucine but also methionine concentration is quite interesting. This is reasonable since the enzyme is located just upstream of the methionine biosynthesis pathway.

Finally, several new functional annotations were made by analysis of regulatory mechanisms and positional clusters of genes. Orthologs of *ygeA* of *E. coli* were predicted to encode branched chain amino acid racemase based on similarity to other racemases and regulation by ILV-attenuator in *P. multocida*. The products of *vatB*, *actX2*, and *actX3* from *P. multocida*, *M. haemolytica*, and *P. filamentus*, respectively, were predicted to catalyze conversion of histamine to 4-β-acetylaminoethyl-imidazole. Three types of predicted histidine transporters are orthologs of *yuiF* and *yvsH* of *B.subtilis*, and *lysQ* of *L. lactis*. They are regulated by candidate histidine attenuators in some bacteria (*HI0325*/*yuiF* in *H. influenzae* and *BC0629*/*yvsH* in *B. cereus*, *lysQ*) and positionally linked to histidine biosynthesis or utilization genes.

This study was partially supported by grants from the Howards Hughes Medical Institute (55000309), the Ludwig Institute of Cancer Research (CRDF RB0-1268), and the Program ''Molecular and Cellular Biology'' of the Russian Academy of Sciences.

Acknowledgements

We are grateful to Andrei Mironov and Dmitry Rodionov for useful discussions, and to Lev Leont'ev for programming assistance.

References

[1] Gollnick, P. and Babitzke, P. (2002) Transcription attenuation. Biochim. Biophys. Acta 1577, 240–250.

[2] 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.

[3] Landick, R., Turnbough, C.L. and Yanovsky, C. (1996) Transcriptional attenuation. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology (Neidhardt, F.C., Ed.), pp. 1263–1286. American Society for Microbiology Press, Washington, DC.

[4] Gelfand, M.S. (1999) Recognition of regulatory sites by genomic comparison. Res. Microbiol. 150, 755–771.

[5] Osterman, A. and Overbeek, R. (2003) Missing genes in metabolic pathways: a comparative genomics approach. Curr. Opin. Chem. Biol. 7, 238–251.

[6] Koonin, E.V. and Galperin, M.Y. (2003) Sequence–Evolution–Function: Computational Approaches in Comparative Genomics. Kluwer Academic Publishers, Boston.

[7] Lyubetskaya, E.V., Leont'ev, L.A., Gelfand, M.S. and Lyubetsky, V.A. (2003) Search for alternative RNA Secondary structures regulating expression of bacterial genes. Mol. Biol. 37, 707–716.

[8] Umbarger, H.E. (1996) Biosynthesis of the branched-chain amino acids. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology (Neidhardt, F.C., Ed.), pp. 442–458. American Society for Microbiology Press, Washington, DC.

[9] Lawther, R.P., Lopes, J.M., Ortuno, M.J. and White, M.C. (1990) Analysis of regulation of the ilvGMEDA operon by using leader-attenuator-galK gene fusions. J. Bacteriol. 172, 2320–2327.

[10] Bartkus, J.M., Tyler, B. and Calvo, J.M. (1991) Transcription attenuation-mediated control of leu operon expression: influence of the number of Leu control codons. J. Bacteriol. 173, 1634–1641.

[11] Wek, R.C. and Hatfield, G.W. (1988) Transcriptional activation at adjacent operators in the divergent-overlapping ilvY and ilvC promoters of *Escherichia coli*. J. Mol. Biol. 203, 643–663.

[12] Rhee, K.Y., Opel, M., Ito, E., Hung, S., Arfin, S.M. and Hatfield, G.W. (1999) Transcriptional coupling between the divergent promoters of a prototypic LysR-type regulatory system, the ilvYC operon of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 96, 14294–14299.

[13] Jafri, S., Chen, S. and Calvo, J.M. (2002) ilvIH operon expression in *Escherichia coli* requires Lrp binding to two distinct regions of DNA. J. Bacteriol. 184, 5293–5300.

[14] Rhee, K.Y., Parekh, B.S. and Hatfield, G.W. (1996) Leucine-responsive regulatory protein DNA interactions in the leader region of the ilvGMEDA operon of *Escherichia coli*. J. Biol. Chem. 271, 26499–26507.

[15] Winkler, M.E. (1996) Biosynthesis of histidine. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology (Neidhardt, F.C., Ed.), pp. 485–505. American Society for Microbiology Press, Washington, DC.

[16] Blasi, F. and Bruni, C.B. (1981) Regulation of the histidine operon: translation-controlled transcription termination a mechanism common to several biosynthetic operons. Curr. Top. Cell. Regul. 19, 1–45.

[17] Patte, J.C. (1996) Biosynthesis of threonine and lysine. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology (Neidhardt, F.C., Ed.), pp. 528–541. American Society for Microbiology Press, Washington, DC.

[18] Green, R.C. (1996) Biosynthesis of methionine. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology (Neidhardt, F.C., Ed.), pp. 542–561. American Society for Microbiology Press, Washington, DC.

[19] Lynn, S.P., Burton, W.S., Donohue, T.J., Gould, R.M., Gumport, R.I. and Gardner, J.F. (1987) Specificity of the attenuation response of the threonine operon of *Escherichia coli* is determined by the threonine and isoleucine codons in the leader transcript. J. Mol. Biol. 194, 59–69.

[20] Lu, Y., Shevtchenko, T.N. and Paulus, H. (1992) Fine-structure mapping of *cis*-acting control sites in the lysC operon of *Bacillus subtilis*. FEMS Microbiol. Lett. 71, 21–27.

[21] Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S. (2003) Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? Nucleic Acids Res. 31, 6748–6757.

[22] Grundy, F.J., Lehman, S.C. and Henkin, T.M. (2003) The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. Proc. Natl. Acad. Sci. USA 100, 12057–12062.

[23] Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S. and Breaker, R.R. (2003) An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev. 17, 2688–2697.

[24] Pittard, A.J. (1996) Biosynthesis of aromatic amino acids. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology (Neidhardt, F.C., Ed.), pp. 458–484. American Society for Microbiology Press, Washington, DC.

[25] Somerville, R. (1992) The Trp repressor, aligand-activated regulatory protein. Prog. Nucleic Acid Res. Mol. Biol. 42, 1–38.

[26] Landick, R., Yanofsky, C., Choo, K. and Phung, L. (1990) Replacement of the *Escherichia coli* trp operon attenuation control codons alters operon expression. J. Mol. Biol. 216, 25–37.

[27] Springer, M., Mayaux, J.F., Fayat, G., Plumbridge, J.A., Graffe, M., Blanquet, S. and Grunberg-Manago, M. (1985) Attenuation control of the *Escherichia coli* phenylalanyl-tRNA synthetase operon. J. Mol. Biol. 181, 467–478.

[28] Gavini, N. and Davidson, B.E. (1991) Regulation of pheA expression by the pheR product in *Escherichia coli* is mediated through attenuation of transcription. J. Biol. Chem. 266, 7750–7753.

[29] Bae, Y.M. and Stauffer, G.V. (1991) Genetic analysis of the attenuator of the *Rhizobium meliloti* trpE(G) gene. J. Bacteriol. 173, 3382–3388.

[30] Yanofsky, C., Konan, K.V. and Sarsero, J.P. (1996) Some novel transcription attenuation mechanisms used by bacteria. Biochimie 78, 1017–1024.

[31] Grundy, F.J. and Henkin, T.M. (2003) The T box and S box transcription termination control systems. Front. Biosci. 8, 20–31.

[32] Panina, E.M., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S. (2001) Regulation of aromatic amino acid biosynthesis in γ-proteobacteria. J. Mol. Microbiol. Biotechnol. 3, 529–543.

[33] Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. (2003) GenBank. Nucleic Acids Res. 31, 23–27.

[34] Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov Jr., E., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., Bhattacharyya, A., Burd, H., Gardner, W., Hanke, P., Kapatral, V., Mikhailova, N., Vasieva, O., Osterman, A., Vonstein, V., Fonstein, M., Ivanova, N. and Kyrpides, N. (2003) The ERGO genome analysis and discovery system. J. Bacteriol. 185, 5673–5684.

[35] Mironov, A.A., Vinokurova, N.P. and Gelfand, M.S. (2000) GenomeExplorer: software for analysis of complete bacterial genomes. Mol. Biol. 34, 222–231.

[36] Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D. and Koonin, E.V. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29, 22–28.

[37] Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.
[38] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and 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.
[39] Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R., Courcelle, E., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Griffith-Jones, S., Haft, D., Hermjakob, H., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Orchard, S., Pagni, M., Peyruc, D., Ponting, C.P., Servant, F. and Sigrist, C.J. (2002) InterPro: an integrated documentation resource for protein families, domains and functional sites. Brief Bioinform. 3, 225–235.
[40] Lyubetskaya, E.V., Leont'ev, L.A., Lyubetsky, V.A. (2003) Detecting of alternative RNA secondary structures in proteobacteria. Information Processes 3, 23–38 (http://www.jip.ru/ – Electronic Scientific Journal) (in Russian).
[41] Gorbunov, K.J., Lyubetskaya, E.V., Lyubetsky, V.A. (2001) Two algorithms for prediction of alternative RNA secondary structures. Information Processes 1, 178–187 (http://www.jip.ru/ – Electronic Scientific Journal) (in Russian).
[42] Popham, D.L. and Setlow, P. (1993) Cloning, nucleotide sequence, and regulation of the *Bacillus subtilis* pbpE operon, which codes for penicillin-binding protein 4 and an apparent amino acid racemase. J. Bacteriol. 175, 2917–2925.
[43] Okada, H., Yohda, M., Giga-Hama, Y., Ueno, Y., Ohdo, S. and Kumagai, H. (1991) Distribution and purification of aspartate racemase in lactic acid bacteria. Biochim. Biophys. Acta 1078, 377–382.
[44] Tarleton, J.C., Malakooti, J. and Ely, B. (1994) Regulation of *Caulobacter crescentus* ilvBN gene expression. J. Bacteriol. 176, 3765–3774.
[45] Patek, M., Krumbach, K., Eggeling, L. and Sahm, H. (1994) Leucine synthesis in *Corynebacterium glutamicum*: enzyme activities, structure of leuA, and effect of leuA inactivation on lysine synthesis. Appl. Environ. Microbiol. 60, 133–140.
[46] Casalone, E., Barberio, C., Cavalieri, D. and Polsinelli, M. (2000) Identification by functional analysis of the gene encoding alpha-isopropylmalate synthase II (LEU9) in *Saccharomyces cerevisiae*. Yeast 16, 539–545.
[47] Olekhnovich, I. and Gussin, G.N. (2001) Effects of mutations in the *Pseudomonas putida* miaA gene: regulation of the trpE and trpGDC operons in *P. putida* by attenuation. J. Bacteriol. 183, 3256–3260.
[48] Delorme, C., Ehrlich, S.D. and Renault, P. (1999) Regulation of expression of the *Lactococcus lactis* histidine operon. J. Bacteriol. 181, 2026–2037.
[49] Chopin, A., Biaudet, V. and Ehrlich, S.D. (1998) Analysis of the *Bacillus subtilis* genome sequence reveals nine new T-box leaders. Mol. Microbiol. 29, 662–664.
[50] Kreimer, D.I., Malak, H., Lakowicz, J.R., Trakhanov, S., Villar, E. and Shnyrov, V.L. (2000) Thermodynamics and dynamics of histidine-binding protein, the water-soluble receptor of histidine permease. Implications for the transport of high and low affinity ligands. Eur. J. Biochem. 67, 4242–4252.
[51] Steffes, C., Ellis, J., Wu, J. and Rosen, B.P. (1992) The lysP gene encodes the lysine-specific permease. J. Bacteriol. 174, 3242–3249.
[52] Patte, J.C., Clepet, C., Bally, M., Borne, F., Mejean, V. and Foglino, M. (1999) ThrH, a homoserine kinase isozyme, with in vivo phosphoserine phosphatase activity in *Pseudomonas aeruginosa*. Microbiology 4, 845–853.
[53] Comparative genomics of the methionine metabolism in gram-positive bacteria: a variety of regulatory systems (2004) Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., Gelfand, M.S., Nucleic Acids Res. (submitted).