DETECTION OF CLASSICAL ATTENUATION IN BACTERIAL GENOMES

Leontiev L.A.*, Shirshin M.A., Lyubetsky V.A.

Institute for Information Transmission Problems, RAS, Moscow, Russia
* Corresponding author: e-mail: leontiev@iitp.ru.

Keywords: *attenuation regulation, threonyl-tRNA synthetase, tryptophan biosynthesis, beta-proteobacteria, branched amino acid biosynthesis, Thermus/Deinococcus group*

Summary

Motivation: Screening of the genomes for signals for patterns of gene expression regulation including attenuation with the leader peptide, is an important task in itself and in the general context of bioinformatics, for instance, in the field of gene annotation. However, development of effective algorithms for mass attenuation detection is is far from complete(for details see, in particular, Lyubetskaya, Molecular Biology 2003; Vitreshak *et al*., 2004). The abstract presents the second part of a collaborative work first described in the conference proceedings in (Lyubetsky *et al*., 2004).

Results: Novel cases of attenuation regulation of the aminoacyl tRNA synthetase, branched amino acid and tryptophan biosynthesis are detected in the beta-proteobacteria, Actinobacteria and Thermus/Deinococcus groups.

Introduction

The "classical" attenuation prototype includes the leader peptide reading frame with a number of regulatory codons and the three hairpins – pause hairpin, antiterminator and terminator with a run of Us (Lyubetskaya *et al*., 2003; Vitreshak *et al*., 2004, in press). We found that such aregulation usually requires three conserved fragments (which we call "three words"), like the bases of the hairpins. Pairing of the first and second word switches on the antiterminator, pairing of the second and third word – the terminator. Supposedly, alternative blocking of the first word by the ribosome determines which one of the alternative conformations is formed.

Methods and Algorithms

Nucleotide sequence data were obtained from the NCBI databases. Putative attenuation structures were detected with an ad hoc algorithm described in (Gorbunov *et al*., 2001; Lyubetskaya *et al*., 2003).

Implementation and Results

1. We predicted attenuation regulation of genes involved in branched aminoacid biosynthesis in the Thermus/Deinococcus group of gram-positive bacteria. In *Deinococcus radiodurans*, attenuation of the gene putatively involved in leucine biosynthesis was found. The regulated gene (having the attenuation structure within the leader segment) named *leuA2* after a homologous gene in *Sinorhizobium meliloti* (homologs are also found in alpha-proteobacteria, Actinobacteria and fungi) encodes the 2-isopropylmalate synthase and is located at the beginning of a 4-gene operon. The other three genes do not have homologs in alpha- and gamma-proteobacteria and *Bacillus subtilis*. In *Thermus thermophilus,* the closest relative of *Deinococcus radiodurans*, three 2-isopropylmalate to synthase isozymes were found but none contains an open reading frame with a set of leucine codons. The attenuation structure is shown below.

Deinococcus radiodurans leuA2. Start codon is atposition 1496949 in the genome. Start, stop and regulatory codons are set in bold, the antiterminator underlined, terminator set in the uppercase.

atgcctgcatacgg**cctgcttcttctt**agc**ctt**cctccggtggag**tga**cggcggacgcagcccgtatcccagCCCCCGGAGGa
ttcagaCCTTCCGGGGGacatttttgt

In addition, the attenuation-regulated operon *ilvBN-x-ilvC* was found in *Deinococcus radiodurans*, where the *ilvBN* genes encode two subunits of acetolactate synthetase, the function of the gene *x* is unknown (its length being 368 bp), and *ilvC* encodes ketol acid reductoisomerase.

Deinococcus radiodurans ilvBN-x-ilvC. Start codon is at position 1531471 in the genome.

atgagcgctggaaac**gtgattgtacttctt**agccacgggggacaccgaggc**taa**gcaggttgtaccggtttcagtcgcACCCCCC
GCCCaaaccaaGGAGGCGGGGGGTttt

Three genes, *ilvGM, ilvBN, ilvIH*, encoding the acetolactate synthase isozymes were found in Enterobacteria, with the first two attenuation regulated, while only one acetolactate synthase was found in *Deinococcus radiodurans*. The same situation is with *Thermus thermophilus* (the genome is not publicly available), which contains only one acetolactate synthase IlvBN. The synthase is located within the *ilvBNC-leuA2-x-leuA2* operon. The genes encode: two acetolactate synthase subunits, ketol-acid reductoisomerase, 2-isopropylmalate synthase LeuA2, ribosome alanine-transferase and 2-isopropylmalate synthase LeuA2' again; *leuA2* and *leuA2'* are 30 % homologous.

2. This article also describes novel attenuators located upstream of the genes of tryptophan and tryptophanyl-tRNA synthetase biosynthesis in some Actinobacteria. *Streptomyces avermitilis* and *Corynebacterium diphtheriae* possess genes *trpE1* and *trpE2* that are homologous to *trpE* of *E. coli*. In *Streptomyces avermitilis*, the *trpE2* gene is a part of an operon uniting genes with unknown function, its homolog in *Corynebacterium diphtheriae* is located in the *x-trpB-trpE2-trpGDC* operon*;* expression of both genes is not attenuation regulated. However, the *trpE1* genes encoding an anthranilate synthase isozyme are separated within the genomes of both organisms, regulated by attenuation.

Streptomyces avermitilis trpE1. Start codon has position 7322414 in the genome.
atgttcgcgcactcgatccagaac**tggtggtgg**accgctcatccggcggcccac**tga**ctgcgcgtacgcaagacttcgcgAAGGCCGCC
CgagGGGCGGCCTTtcgtgtt
Corynebacterium diphtheriae trpE1. Start codon has position 2456514 in the genome.
atgaatgcacataac**tggtggtgg**cgcgct**taa**ccgcggggccgttttcacgcattcatttcaacAGGCTCgc
CTTGTccaACAAGCAGCGGGCCTttttgtta
Also in *Streptomyces avermitilis*, the *trpS2* gene encoding the tryptophan-tRNA synthetase is probably regulated by attenuation each.
Streptomyces avermitilis trpS. Start codon is at 5758647 in the genome.
atgactacgcgtacgtgtacccagcag**tggtgg**gccgcc**tga**cggcggccgtacacacgtatgtactcAACGGCCGCCG
cctCGGCGGCCGTTctcgttt

This organism contains two tryptophan-tRNA synthase isozymes encoded by *trpS* and *trpS2,* both genes are homologous to *trpS* of *Escherichia coli* and *Corynebacterium duphtheriae*. Notably, the *trpS* genes in *Escherichia coli* and *Corynebacterium diphtheriae* do not have paralogs and are probably not regulated by attenuation.

3. Putative attenuators of threonyl-tRNA synthetase were predicted in beta-proteobacteria, namely, in *Bordetella* spp. (*Bordetella bronchiseptica*, *Bordetella parapertussis*, *Bordetella pertussis*), *Ralstonia* spp. (*Ralstonia metallidurans*, *Ralstonia solanacearum*), *Chromobacterium Vilaceum,* and in *Methylococcus capsulatus,* which is tentatively classified within gamma-proteobacteria. In *Bacillus subtilis*, two isozymes of the threonyl tRNA synthase were found, *thrS* and *thrZ*, which are 59 % homologous. The first, *thrS*, is not regulated by attenuation, and the second one, *thrZ*, contains a threonine-dependent attenuator.

Bacillus subtilis thrZ. Start codon is at 3855364 in the genome.

atgctgcgttacagcaccgagccgacaaca**tag**tatttgtcgggaacttgggtggaaccacgggttaatcACACACTCGTCCC
TATCTGCGGGACGGGTGTGTtttttta

In *Escherichia coli*, only one *thrS* gene was found. The gene encodes threonyl-tRNA synthase

and is not attenuation regulated. Attenuation structures in *Bordetella* spp. are completely identical in sequence. The one from *Bordetella pertussis* is shown below.

Bordetella pertussis thrZ. Start is at1574598 in the genome.

atgctgctgcgcccgacacgaactacgaccggaatcttccgcactca
gtcgcgt**taa**tgatttcatgcgtcgggttggtatcgcgtcattgcggtaaatactcaacaaggcaccaGACAAAACGCGGC
caggcaGCCGCGTTTTTCgtttcc

Attenuation structures in *Ralstonia solanacearum*, *Chromobacterium vilaceum* and *Methylococcus capsulatus* are also shown below.

Ralstonia metallidurans Reut_370 (starts at 108811)

atgagcaagacaactcgaactactaccgctggt**tag**cgacagtagtcga
ggtgcgccttcgacccgtagtgtgtgatacggggaaacacagaaaaACGCGGCC
ttGGCCGCGTttttttcgtc

Ralstonia solanacearum (starts at 1692246)

atgatccaggcaccgcgaactaccaccgcttcggagcgacag**tag**tcaaggt
gcggtttcgccctttgcggtctgcccaagcggcagttcagcgcaaatgaaaACGC
GGCCttGGCCGCGTttttttcgtccgcgt

Chromobacterium Vilaceum ATCC 12472 (starts at 1420122)

atggtgggggtcgttggttcgaatccaatcgtgcctaccaaattaaatgcatggagttgccaagtgaaatccttgcgaactaccacaacccgacatactcacgaaa
gc**tga**gccttgcgggttggaccatgcgaaaaAGAAGTGCGGCTcaGGCCGCACTTTTtttttaccctt

Methylococcus capsulatus str. Bath (starts at 729164)

atgcctccttcttgtcgaattacgacaggcgcg**tag**ctcagt
tggttagagcaccaccttgacatggtgggggtcgttggttcgagtccaatcgcgcctaccagatatcccagGAAGCATCGCCa
GGCGATGCTTTtttgtttggggg

Discussion

Novel cases of attenuation regulation of biosynthesis of several aminoacyl-tRNA synthetases, branched amino-acids and tryptophan are found in bacteria *Deinococcus radiodurans, Thermus Thermophilus, Streptomyces avermitilis, Corynebacterium diphtheriae* and the groups of beta-proteobacteria Bordetella spp, Ralstonia spp, *Chromobacterium vilaceum, Methylococcus capsulatus*. The regulation patterns are partly described above and dealt with in in more detail in the presentation.

The conventional attenuation is a universal mechanism of gene expression regulation in terms of both its wide occurrence across the organismal diversity (it is reported for many bacterial taxa from Bacteroidetes /Chlorobi, Firmicutes, Thermus/Deinococcus, Actinobacteria, *etc.*) and functional diversity of the regulated operons (amino acid, animoacyl-tRNA synthetases, transporters biosynthesis). Interestingly, the threonyl-tRNA synthetase regulation in *Methylococcus capsulatus* putatively classified within the gamma-proteobacteria is similar to the analogous mechanism in beta-proteobacteria, while within the other gamma-proteobacteria it is not reported.

Acknowledgments

We are grateful to M.S. Gelfand, D.A. Rodionov, A.G. Vitreshak and A.V. Seliverstov for helpful discussions, as well as to V.V. Zubov for providing us with some biological data.

References

Gorbunov K.Yu., Lyubetskaya E.V., Lyubetsky V.A. On two algorithms of detection of alternative elements of RNA secondary structure // Informational processes. 2001. V. 1(2). P. 178–187.

Lyubetskaya E.V., Leontiev L.A., Gelfand M.S., Lyubetsky V.A. Search for alternative secondary tRNA structures, regulating bacterial gene expression // Mol. Biol. 2003. V. 37, N 5. P. 834–842.

Lyubetskaya E.V., Leontiev L.A., Lyubetsky V.A. Alternative secondary structure detection in a group of gamma-bacteria // Informational processes. 2003. V. 3(1). P. 23–38.

Lyubetsky V.A., Seliverstov A.V. Amino Acid Biosynthesis Attenuation in Bacteria // This conference. 2004.

Vitreshak A.G., Lyubetskaya E.V., Shirshin M.A., Gelfand M.S., Lyubetsky V.A. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis // FEMS Microbiology Letters. Accepted, 2004. 39 p.