# AMINO ACID BIOSYNTHESIS ATTENUATION IN BACTERIA

***Lyubetsky V.A., Seliverstov A.V.* ***

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

**Keywords:** *amino acid biosynthesis, attenuation regulation*

## Summary

*Motivation:* Discovery and analysis of the "conventional" attenuation regulation in bacteria is currently an active field of research within the general framework of studying RNA-based regulation strategies, refer to (Vitreschak *et al*., in press) for details. However, finding alternative regulation systems is still a long way even in bacteria. Further, understanding the attenuation regulation machinery is crucial for developing algorithmic tools of mass attenuation detection and deriving a descriptive attenuation model. Identifying relevant attenuation characteristics to study is by itself a separate task.

*Results:* Putative attenuator structures in amino acid biosynthesis in Actinobacteria and *Staphylococcus aureus* are identified. Analysis of biosynthesis attenuation regulation in a wide range of proteobacteria and Gram-positive bacteria provided estimates of its characteristics.

## Introduction

It is to be kept in mind that the "conventional" attenuation regulation of amino acid biosynthesis (this study operates with branched amino acids and leucyl-tRNA synthetase) implies presence of a leader peptide bearing regulatory codons (in fact, not only those encoding the amino acid), a terminator, antiterminator, a pause hairpin and a U-motif. Lyubetskaya *et al*. (2003) hypothesized that the hairpin formation requires a unique triplet word pattern, which was effectively implemented in the LLLM algorithm for mass detection of attenuation regulation (Gorbunov *et al*., 2001; for the search performance see Lyubetskaya *et al*., 2003; Vitreschak *et al*., in press).

## Methods and Algorithms

All nucleotide sequences of leader regions as well as the gene annotations are obtained from NCBI. Conservative anchor motifs for use in multiple alignment were detected by our algorithm (Lyubetsky, Seliverstov, 2003). This algorithm involves the finding cliques in multipartite graph. It requires only polynomial time for computing a set of similar words in each nucleotide sequence.

## Results

***Actinobacteria.*** In many actinobacteria genes *ilvB*, *ilvN* (or *ilvH*) and *ilvC* comprise a single operon. The Table 1 shows the *ilvB*-containing operons' putative leader peptides, the operon type (second column) and the leader peptides' first nucleotide position as according to the NCBI nomenclature (third column).

**Table 1.** Leader peptides

| | | |
|---|---|---|
| *Corynebacterium diphtheriae* | *ilvBHC* | 1081747 |
| Met Asn **Ile Ile** Arg **Leu Val Val Ile** Thr Thr Arg Arg **Leu** Pro | | |
| *Corynebacterium efficiens* YS-314 | *ilvBHC* | 1432212 |
| Met Thr Ser **Ile** Arg Pro **Val Val Ile Val** Ala Ala Arg Arg **Leu** Pro | | |
| *Corynebacterium glutamicum* ATCC 13032 | *ilvBHC* | 1337840 |
| Met Thr **Ile Ile** Arg **Leu Val Val Val** Thr Ala Arg Arg **Leu** Pro | | |
| *Mycobacterium tuberculosis* H37Rv | *ilvBNC* | 3363125 |

| | | |
|---|---|---|
| Met **Leu Val Val Ile** Gly Arg Arg **Val** Gly Ala | | |
| *Mycobacterium bovis* subsp. bovis AF2122/97 | *ilvB-serA1* | 3319743 |
| Met **Leu Val Val Ile** Gly Arg Arg **Val** Gly Ala | | |
| *Mycobacterium leprae* | *ilvBNC* | 2046378 |
| Met **Leu Val Val Ile** Cys Gln Arg **Val** Gly Gly | | |
| *Mycobacterium avium* subsp. paratuberculosis str. k10 | *ilvB1NC* | 3381051 |
| Met **Leu Val Val Ile** Arg Arg **Val** Gly Ala | | |
| *Mycobacterium marinum* | *ilvB* | 166742 |
| Met Asp Thr Ala Gly Thr Pro Gly Lys **Leu Val Val Leu** Gly Arg Arg **Val Val** Ala | | |
| *Streptomyces avermitilis* MA-4680 | *ilvBNC* | 3356481 |
| Met Arg Thr Arg **Ile Leu Val Leu** Gly Lys Arg **Val** Gly | | |
| *Streptomyces coelicolor* A3(2) | *ilvBNC* | 6002909 |
| Met Arg Thr Arg **Ile Leu Val Leu** Gly Lys Arg **Val** Gly | | |

Our alignment reveals that terminator hairpins and their preceding motifs are highly conservative in the organisms studied. Hereafter, the terminator half-stems are set in uppercase and the right-hand antiternimator parts are underlined:

*C. diphtheriae* cgaaaagcGCCCTCGaCAGCAccacacaTGCTGagCGGGGGCtttccttat

*C. efficiens* caagcGCCCTCGACAGTACccaccacaGTGCTGttTCGAGGGCtttgttgt

*C. glutamicum* caagcGCCCTCGaCAACACTcaccacAGTGTTGgaaCGAGGGCtttcttgtt

*M. tuberculosis* ccaacgcgACCCTCGtgCAGCagctgaGCTGgCGAGGGTtttttctt

*M. bovis* ccaacgcgACCCTCGtgCAGCagctgaGCTGgCGAGGGTtttttctt

*M. leprae* ccaacgcgcAACCCTCGtgCAGCTagtcAGCTGtCGAGGGTTttttgtt

*M. avium* ccaacgcgcAACCCTCGtgCAGCacaaGCTGtCGGGGGTTttttgtt

*M. marinum* ccaacgcgcAACCCTCGTgCAGCagctgaGCTGACGGGGGTTttttgtt

*S. avermitilis* cccggcgcgctCCCCTCGctTGCCtcacGGCACGAGGGGttttttgtt

*S. coelicolor* ccgacgcgctCCCCTCGctTGCCttacGGCACGAGGGGttttttgtt

In two actinobacteria, *Streptomyces avermitilis* and *Streptomyces coelicolor*, putative transcription attenuation regulation was found for a leucyl-tRNA synthetase gene *leuS* ortholog. The leader peptide: Met Arg Ala Val Arg **Leu Leu Leu** Ser Glu Pro Arg. Terminator hairpins are in uppercase, antiterminators underlined.

*S. avermitilis* (first nucleotide 6661741)
atgcgtgccgtacgc**cttctgctt**agcgagcc
gcgc**tga**tcagcccagaccactgacgattcgtggtcggaatcggcgcggcgtCCCCTCctgtgcGAGGGGttttttcatt

*S. coelicolor* (first nucleotide 2778624)
atgcgtgccgtacgc**cttctgctt**agcgagcc
gcgc**tga**tcagtcccgaccccggtcgtagtccggtggccggaatcggcgcggcgTCCCCTCctgtgcGAGGGGAtttttcatt

***Staphylococcus aureus***. The leader region of gene *ilvD,* which encodes dihydroxy-acid dehydratase in Gram-positive bacterium *S. aureus* contains a leader peptide preceded by a GA-rich SD region, a terminator hairpin with a U-rich motif and an antiterminator hairpin. The leader peptide possesses leucine and isoleucine codon strings:

Met **Leu** Asn Gln Tyr Thr Glu His Gln Pro Thr Thr Ser Asn **Ile Ile Ile Leu Leu** Tyr Ser **Leu** Gly **Leu** Glu Arg.

There are detected hairpins:
atg**ctt**aatcaatatactgaacatcaaccgacaacttcaaat**attattattttatta**tactct**tta**gga
**ctc**gaacgt**tag**taaatatttactaaacgctttaagtcctatttctgtttgaatgggacttgtAAACGTCCCAATAaTATTGGGACGTTTttttt

The first nucleotide position: N315 – 2097353; Mu50 – 2173855; MW2 – 2125745.

**Table 2.** Attenuation parameters

| Bacteria | Operon | SU | L | G | C | 100G/(G+C) |
|---|---|---|---|---|---|---|
| Actinobacteria | | | | | | |
| *Corynebacterium diphtheriae* | *ilvBHC* | 62 | 7 | 8 | 3 | 88 |
| *Corynebacterium efficiens* | *ilvBHC* | 69 | 8 | 7 | 3 | 70 |
| *Corynebacterium glutamicum* | *ilvBHC* | 67 | 6 | 8 | 2 | 80 |
| *Mycobacterium tuberculosis* | *ilvBNC* | 57 | 6 | 7 | 2 | 77 |
| *Mycobacterium bovis* | *ilvB-serA1* | 57 | 6 | 7 | 2 | 77 |
| *Mycobacterium leprae* | *ilvBNC* | 74 | 4 | 6 | 2 | 75 |
| *Mycobacterium avium* | *ilvB* | 72 | 4 | 7 | 2 | 77 |
| *Mycobacterium marinum* | *ilvB* | 59 | 6 | 7 | 2 | 77 |
| *Streptomyces avermitilis* | *ilvBNC* | 84 | 4 | 7 | 2 | 77 |
| *Streptomyces coelicolor* | *ilvBNC* | 84 | 4 | 7 | 2 | 77 |
| *Streptomyces avermitilis* | *leuS* | 66 | 6 | 5 | 0 | 100 |
| *Streptomyces coelicolor* | *leuS* | 70 | 6 | 5 | 0 | 100 |
| *Corynebacterium diphtheriae* | *trpE1* | 64 | 3 | 5 | 5 | 50 |
| *Streptomyces avermitilis* | *trpE1* | 47 | 3 | 5 | 3 | 62 |
| *Streptomyces avermitilis* | *trpS2* | 52 | 3 | 5 | 4 | 55 |
| Staphylococcus | | | | | | |
| *Staphylococcus aureus* | *ilvD* | 78 | 1 | 4 | 1 | 80 |
| Other | | | | | | |
| *Deinococcus radiodurans* | *leuA2* | 59 | 7 | 5 | 4 | 55 |
| *Deinococcus radiodurans* | *ilvBN-x-C* | 57 | 7 | 10 | 1 | 91 |
| *Thermus Thermophilus* | *ilvBNC* | 45 | 5 | 5 | 2 | 71 |
| *Bordetella* | *thrS* | 86 | 6 | 3 | 3 | 50 |
| *Ralstonia* | | 78 | 2 | 4 | 3 | 57 |
| *Chromobacterium Vilaceum* | | 51 | 2 | 3 | 4 | 43 |
| *Methylococcus capsulatus* | | 101 | 1 | 4 | 2 | 66 |

**Table 3.** Earlier results for proteobacteria

| | Operon | SU | L | G | C | 100G/(G+C) | AS |
|---|---|---|---|---|---|---|---|
| alpha | *ilvIH* | 51-55 | 4-7 | 2-5 | 1-3 | 40-66 | |
| | *trp(E/G)* | 52-72 | 3-10 | 4-6 | 2-5 | 44-66 | |
| gamma | ilvBN | 53-57 | 4-6 | 6-7 | 3 | 66-70 | 6 ..7 |
| | ilvGMEDA | 37-64 | 4-6 | 5-8 | 0-3 | 66-100 | 7 .. 31 |
| | leuABCD | 42-69 | 3-7 | 4-5 | 1-3 | 64-83 | 5 .. 33 |
| | thrABC | 46-62 | 3-8 | 3-7 | 1-3 | 50-88 | -2 .. 22 |
| | his | 90-113 | 3-7 | 2-5 | 1-4 | 50-83 | -6 .. 22 |
| | trp | 44-73 | 4-8 | 3-4 | 1-2 | 60-80 | -8 .. -2 |
| | pheA | 61-72 | 3-5 | 4-6 | 1-2 | 71-86 | -8 .. -6 |
| | pheST | 68-69 | 3-6 | 4-5 | 1 | 80-83 | 5 .. 33 |

The Tables 2 and 3 show some characteristics of transcription attenuation regulation. Data for proteobacteria are originally from (Vitreschak *et al*., in press). The third column contains distances SU between the initial position of the leader peptide stop codon and the beginning of the U-rich terminator hairpin region. The loop size of newly predicted terminators does not exceed 8, which well conforms to the known cases. The fifth and sixth columns contain the amount of G and C bases in the right half-stem of the terminator (preceding the poly-U). The distance AS between the antiterminator left half-stem and the stop codon varies between -8 (stop codon to the left of the antiterminator) and 33 (stop codon in the middle of the antiterminator loop).

## Discussion

For *ilvB*-containing operons the distance SU is larger than in known proteobacteria. However, in operons *pheA*, *pheST* and *trp(E/G)* it is even larger and reaches 113 bases in *hisGDCBHAFI* operons. This parameter is a characteristic of the antiterminator structure properties. When the terminator hairpin is enough GC-rich, the proportion of Gs in its right half-stem is higher than that of Cs. The average ratio G/(G+C) = 2/3. The exception are some proteobacteria with a very short terminator containing nearly equal amounts of G and C (for instance, the right half-stem of the operon *ilvIH* terminator in *Rhodopseudomonas palustris* has a higher C content) and also low-GC bacteria. In the *ilvBNC* operon of actinobacteria predicted terminators possess a longer hairpin with the relative G content close to that in gamma-proteobacteria. Predictions in actinobacteria and other Gram-positive bacteria conform well to previous results. A stop codon can not be situated considerably far to the left from the antiterminator (rather, from the nucleotides complementary to a terminator hairpin region). The assumption AS > -9 seems to be strict. The number of regulatory codons in the leader peptide strongly correlates with the encoded amino acid. For tryptophan, a duplet or triplet of adjacent codons suffice. The *ilv* and *thr* operons involved in biosynthesis of several amino acids have leader peptides with numerous regulatory codons (14 codons preceding *ilvGMEDA* in *E. coli*). However, the *ilvIH* operon's leader peptide in alpha-proteobacteria contains between 3 (*Caulobacter crescentus*) and 6 regulatory codons. Predictions for *ilv* in actinobacteria and *S.aureus* contain at least 5 regulatory codons, which is congruent with evidence.

## Acknowledgements

We wish to thank Dr. M.S. Gelfand for critical comments on this research.

## References

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

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

Lyubetsky V.A., Seliverstov A.V. Selected algorithms related with finite groups // Information processes. 2003. V. 3(1). P. 39–46.

Vitreschak A.G., Lyubetskaya E.V., Shirsin M.A., Gelfand M.S., Lyubetsky V.A. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis // FEMS Microbiology Letters. 2004, (in press).