BIOINFORMATICS APPROACH TO ANALYSIS OF REGULATION OF AROMATIC AMINO ACIDS BIOSYNTHESIS IN *BACILLUS/CLOSTRIDIUM* GROUP

[1] Panina E., [1,2] Vitrehschak A., [1] Mironov A., [1] Gelfand M.

[1] Branch of Corporation Integrated Genomics, Inc., postbox 348, 117333, Moscow, Russia
[2] Institute of Problems for Information Transmission, RAS
e-mail: katya@ekpanina.mccme.ru
*Corresponding author

Key words: *aromatic amino acids, comparative genomics, T-box, TRAP, transcription, regulation, ABC transporter*

Resume

Motivation: While regulation of aromatic amino acids biosynthesis (AAAB) has been intensely studied in *Bacillus subtilis*, little is known about the mechanisms of regulation in other members of the *Bacillus*/*Clostridium* group. Since most species in this group are dangerous human pathogens, e.g. *Bacillus anthracis* and *Staphylococcus aureus*, the theoretical research in this area is highly important.
Results: We have applied the comparative genomics approach to analysis of regulatory patterns involved in AAAB in the *Bacillus/Clostridium* group. We demonstrate the variability of DNA and RNA regulation of orthologous genes in different species. We describe a new type of transcriptional regulation of DAHP synthase and shikimate kinase genes in the *Streptococcus* and *Lactococcus* species, and new candidate T-boxes upstream of AAAB genes in the analyzed genomes. Finally, we identify a candidate tryptophan transporter in the *Streptococcus*, *Lactococcus*, *Enterococcus,* and *Desulfitobacterium* species.

Introduction

Biosynthesis of three aromatic amino acids starts with the common pathway leading from phosphoenolpyruvate (PEP) and erythrose 4-phosphate (E4P) through 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) and shikimate to the chorismic acid (genes *aroA, aroB, aroC, aroI, aroD, aroE,* and *aroF* in *Bacillus subtilis*). Then, the pathway divides into the terminal pathways, specific for each aromatic amino acid (genes *trpE, trpG, trpD, trpC, trpF, trpB,* and *trpA* for the tryptophan production; *aroA, pheA, pheB, aroH, tyrA, hisC,* and *aspB* for the phenylalanine and tyrosine production in *Bacillus subtilis*).
In gram-positive bacteria, no transcriptional regulation of AAAB has yet been experimentally discovered. However, Terai et al. (2001) have identified PCEs (phylogenetically conserved elements) upstream of *aroA* genes in *B. subtilis* and *Bacillus halodurans*, and upstream of *aroF* genes in the *B. subtilis* and *Bacillus stearothermophilus*, which might play a role in the transcriptional regulation of AAAB in *Bacillus* species. The RNA regulation of this pathway in gram-positive bacteria involves the RNA-binding protein TRAP that regulates transcription and translation of the *trpEDCFBA* operon and translation of the *trpG* and *yhaG* genes in *B. subtilis*, the latter encoding a candidate tryptophan-specific permease (Bobitzke, Gollnick, 2001). The other type of the RNA-level regulation is presented by T-boxes that regulate transcription of the *trpEGDCFBA* operon in *Lactococcus lactis* (Fig. 1).
We have previously applied the comparative genomics approach to the analysis of DNA- and RNA-level regulation of AAAB in γ-proteobacteria (Panina et al., 2000). Here, we apply the same approach to the analysis of regulatory patterns involved in this pathway in gram-positive bacteria of the *Bacillus/Clostridium* group: *Bacillus*, *Clostridium, Streptococcus, Enterococcus, Lactococcus, Staphylococcus, Lysteria,* and *Desulfitobacterium* species.

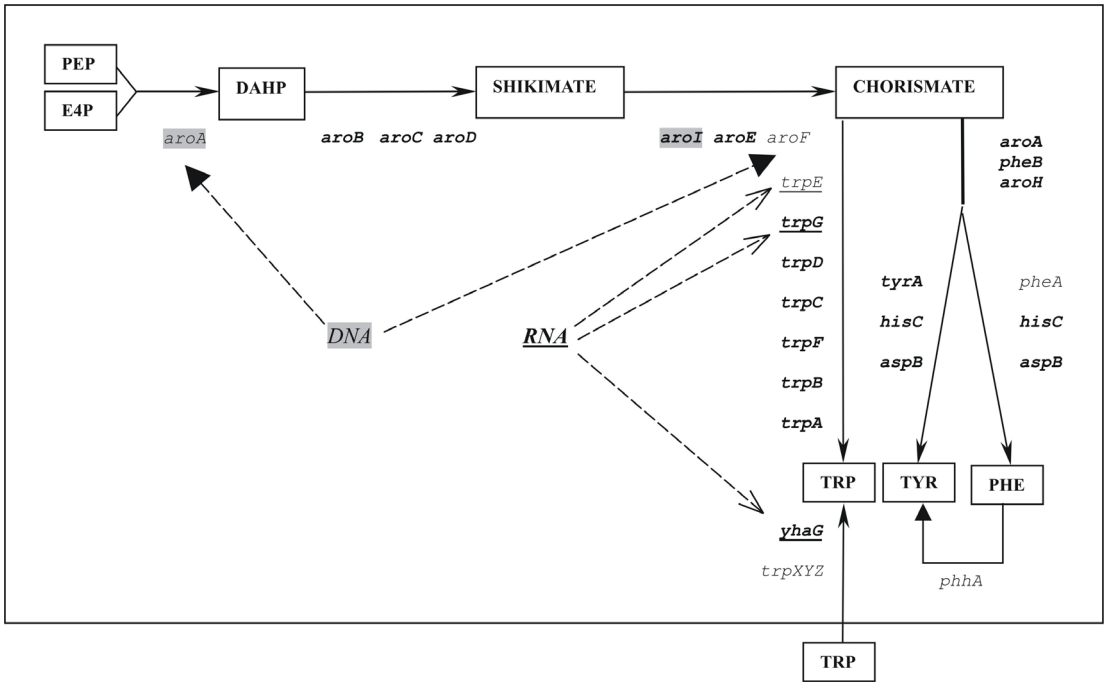

Fig. Genes encoding the enzymes of the aromatic amino acids biosynthesis pathway, their regulation, and the transporters for tryptophan (*yhaG* is a known transporter, and *trpXYZ* is predicted to be a tryptophan transporter in this study). The known regulation is shown by dotted lines: filled arrows, DNA-level regulation and PCEs; empty arrows, RNA level regulation, TRAP (underlined), and T-boxes (bold). Candidate regulation found in this study: shaded, new type of transcription regulation and bold, new T-boxes.

Materials and Methods

Complete genome sequences of *Bacillus subtilis, B. halodurans, Streptococcus pneumoniae, Lactococcus lactis, Enterococcus faecalis, Streptococcus pyogenes, Clostridium acetobutilicum, Staphylococcus aureus,* and Listeria monocytogenes were downloaded from GenBank (Benson et al., 2000). Partially sequenced genomes of *Bacillus stearothermophilus, Streptococcus mutans, Clostridium difficile,* and *Desulfitobacterium halfniense* were extracted from the ERGO Database (http://wit.mcs.anl.gov/WIT2/). Partially sequenced genome of *Enterococcus faecium* was obtained from the DOE Joint Genome Institute site (http://www.jgi.doe.gov); and partially sequenced genome of *Bacillus anthracis* was obtained from the WWW site of the Institute for Genomic Research (http://www.tigr.org).

Profiles for signal recognition were constructed as described in (Panina et al., 2001). Positional nucleotide weights in these profiles are defined as

$W(b,k) = \log[N(b,k) + 0.5] - 0.25\Sigma_{i=A,C,G,T}\log[N(i,k) + 0.5]$,

where N(b,k) is the count of nucleotide b at position k. The score of the candidate site is calculated as the sum of the respective positional nucleotide weights:

$Z(b_1 \ldots b_L) = \Sigma_{k=1\ldots L} W(b_k,k)$, where k is the length of the site.

Genomic analyses (protein similarity searches using Smith-Waterman algorithm, analysis of orthology, and identification of candidate signals in the genome sequences) were done using GenomeExplorer (Mironov et al., 2000). Searches for RNA secondary structure sites were performed using RNApattern.

Results and Discussions

The pathway: genes and operons. While the backbone of AAAB pathway is conserved in most bacterial species, we have identified some steps that vary within the analyzed group. First, the complete genomes of *S. pyogenes* and *E. faecalis* lack genes for terminal tryptophan pathway. Second, in *S. pyogenes* there are no homologs of *pheA* and *tyrA* genes from the terminal phenylalanine and tyrosine pathways, respectively. Third, in *S. pneumoniae, S. mutans,* and *L. lactis,* there are no homologs of the DAHP synthase gene *aroA* of *B. subtilis*, while there are two genes homologous to DAHP synthases from gram-negative bacteria. Next, in *B. anthracis* and *D. halfniense*, there is a homolog of the *phhA* gene previously identified only in a number of Proteobacteria and eukaryotes. PhhA catalyzes the conversion of phenylalanine to tyrosine. Finally, the *Bacillus, Streptococcus,* and *Clostridium* genomes, excluding only *B. anthracis*, have one copy of *trpG* gene that functions both in the tryptophan and folate biosynthesis, whereas *L. lactis, S. aureus, B. anthracis,* and *D. halfniense* have two paralogous copies of this gene.

The operon structure of AAAB genes varies significantly. The only conserved feature is the *trpE*(*G*)*DCFBA* operon, which is either absent or present as a whole. The only exception is the *trpG* gene that lies either in the *trp* operon (in

S. pneumoniae, S. mutans, and *C. acetobutilicum*), or in the folate biosynthesis operon (in *B. subtilis, B. halodurans, B. stearothermophilus, C. difficile,* and *S. pyogenes*). In *L. lactis, S. aureus, B. anthracis,* and *D. halfniense*, where there are two copies of *trpG* gene, one copy lies in the *trp* operon, whereas the other one is co-localized with the folate biosynthesis genes. Thus, we propose that the duplicated enzymes have acquired narrow specificity for tryptophan (RLLX01504, RSA03401, RZC03347, RDHA05110) and folate (RLLX01346, RSA02493, RZC04175, RDHA04984) production, respectively.

DNA-level regulation. Pairs of DAHP-synthase genes of *S. pneumoniae*, *S. mutans,* and *L. lactis,* encoding homologs to gram-negative, rather than gram-positive enzymes, form operons in *S. pneumoniae*, *S. mutans*, but are located separately in *L. lactis.* We have found a conserved 14-bp sequence ATGGAGGCANATAA upstream of the DAHP synthase operons in *S. pneumoniae* and *S. mutans*, and upstream of both DAHP synthases genes in *L. lactis.* Moreover, a similar sequence was found in the upstream regions of the shikimate kinase genes in all the three species. Notably, the reactions catalyzed by shikimate kinase and DAHP synthase are the only two irreversible steps within the common pathway of AAAB, and only these genes of the common pathway are regulated at the transcriptional level in γ-proteobacteria. Thus, we propose that the new conserved sequence plays a role in transcriptional regulation of DAHP synthase and shikimate kinase genes in *Streptococcus* and *L. lactis* genomes.

We have constructed the profile based on the PCEs described in (Terai et al., 2001). Using this profile we have found one more candidate site ACTTAAccaCGTT upstream of the *aroF* gene in *B. halodurans.*

RNA-level regulation. A number of T-boxes were found upstream of genes involved in AAAB. In particular, tyrosine-specific T-boxes were found upstream of the *aroA*, *aroF,* and *phhA* genes in *B. anthracis*; tryptophan-specific T-boxes were found upstream of the *trp* operons in *B. anthracis*, *S. pneumoniae, S. mutans, L. lactis, C. acetobutilicum, S. aureus,* and *L. monocytogenes.* We have also found a phenylalanine-specific T-box upstream of the *pheA* gene in *D. halfniense.*

Candidate TRAP-binding sites were found upstream of the *trp* operons and *trpG* genes in *B. halodurans* and *B. stearothermophilus*.

Interchange of regulatory systems. So far, there seem to be four types of regulation of AAAB in *Bacillus/Clostridium* group. The most general is the T-box-dependent transcriptional regulation, which is present in all the studied species. Another type of the RNA-dependent transcriptional regulation, TRAP-mediated regulation, is unique to the *Bacillus* group except for *B. anthracis*, which lacks the TRAP protein. In *B. subtilis*, *B. halodurans,* and *B. stearothermophilus*, TRAP regulates transcription of the *trp* operon, which is regulated by tryptophan-specific T-boxes in all the other species. The third type of regulation, PCEs, is also specific of *B. subtilis*, *B. halodurans* and *B. stearothermophilus*, where it appears to regulate the transcription of DAHP synthase and chorismate synthase genes. In *B. anthracis,* the same genes are regulated by tyrosine-specific T-boxes, while in *S. pneumoniae*, *S. mutans,* and *L. lactis,* DAHP synthases as well as shikimate kinases genes are under the control of the fourth type of transcriptional regulation identified in this study.

Transporters of aromatic amino acids. The only known tryptophan transporter in the *Bacillus/Clostridium* group is YhaG of *B. subtilis*, whose translation is regulated by TRAP protein. We have found orthologs of the *yhaG* gene in *B. stearothermophilus*, *C. acetobutilicum,* and *C. difficile*; however, no homologs of *yhaG* could be observed in the genomes of *E. faecalis* and *S. pyogenes,* which lack the tryptophan biosynthesis pathway, and thus, should transport tryptophan from the environment. We have identified tryptophan-specific T-boxes upstream of the *yhaG* orthologs in both *Clostridium* species; in *B. stearothermophilus,* the upstream region of this gene is not sequenced yet.

We have found a new candidate tryptophan ABC transporter, named *trpXYZ*, in the genomes of *S. pneumoniae*, *S. mutans*, *S. pyogenes, S. equi*, *E. faecalis*, *E. faecium, B. stearothermophilus, D. halfniense, B. cepacia,* and *M. loti* (the last two are α-proteobacteria). The genes in *S. pneumoniae genome* are *SP1069, SP1070, and SP1071*. *trpXYZ* is presented in three copies in the genome of *D. halfniense*, and two of them have tryptophan-specific T-boxes in the upstream regions. Besides, *trpXYZ* is preceded by a tryptophan-specific T-box in *S. pneumoniae.* Moreover, *trpXYZ* is co-localized with the *aroD* gene in *E. faecium,* and with gene encoding enzymes of the tryptophan degradation kynurenine pathway in *M. loti.* These observations allow us to ascribe the tryptophan specificity to this transporter.

Acknowledgements

We are grateful to Dmitry Rodionov for useful discussions. This study was partially supported by grants from INTAS (99-1476) and HHMI (55000309).

References

1. Bobitzke P., Gollnick P. (2001). Posttranscriptional initiation control of tryptophan metabolism in *Bacillus subtilis* by the trp *RNA*-binding attenuation protein (TRAP), anti-TRAP, and RNA structure. J. Bacteriol. 183, 5795-5802.
2. Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Rapp B.A., Wheeler D.L. (2000). GenBank. Nucl. Acids Res. 28, 15-18.
3. Mironov A.A., Vinokurova N.P., Gelfand M.S. (2000). GenomeExplorer: software for analysis of complete bacterial genomes. Mol. Biol. 34, 222-231.
4. Panina E.M., Vitreschak A.G., Mironov A.A, Gelfand M.S. (2001). Regulation of aromatic amino acid biosynthesis in gamma-proteobateria. J. Mol. Microbiol. Biotechnol. 3, 529-543.
5. Terai G., Takagi T., Nakai K. (2001). Prediction of co-regulated genes in *Bacillus subtilis* on the basis of upstream elements conserved across three closely related species. Genome Biol. 2, research 0048.1–0048.12.