New insights into regulation of the tryptophan biosynthetic operon in Gram-positive bacteria

A. Gutierrez-Preciado[1], R.A. Jensen[2], C. Yanofsky[3] and E. Merino[1]

[1]Department of Molecular Microbiology, Instituto de Biotecnologia, Universidad Nacional Autonoma de Mexico, Cuernavaca, Morelos 62271, Mexico
[2]Emerson Hall, Bldg 981, PO Box 110700, University of Florida, Gainesville, FL 32611, USA
[3]Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020, USA

The tryptophan operon of *Bacillus subtilis* serves as an excellent model for investigating transcription regulation in Gram-positive bacteria. In this article, we extend this knowledge by analyzing the predicted regulatory regions in the *trp* operons of other fully sequenced Gram-positive bacteria. Interestingly, it appears that in eight of the organisms examined, transcription of the *trp* operon appears to be regulated by tandem T-box elements. These regulatory elements have recently been described in the *trp* operons of two bacterial species. Single T-box elements are commonly found in Gram-positive bacteria in operons encoding aminoacyl tRNA synthetases and proteins performing other functions. Different regulatory mechanisms appear to be associated with variations of *trp* gene organization within the *trp* operon.

Introduction

The cost of synthesizing an essential metabolite is often appreciable; therefore, it is appropriate that each organism tightly regulates the expression of genes in each metabolic pathway. This is particularly important for the genes encoding proteins involved in tryptophan biosynthesis, one of the most 'expensive' biosynthetic pathways. A comprehensive regulatory model for the genes involved in the tryptophan biosynthetic pathway has been proposed for the Gram-positive bacterium *Bacillus subtilis* (reviewed in Refs [1] and [2]). In this organism, a six gene *trp* operon is located within a 12-gene aromatic supraoperon. Three genes precede and three genes follow the *trp* operon (Figure 1) [2,3]. These six additional genes encode proteins that function in the common aromatic pathway in phenylalanine or tyrosine biosynthesis. Two promoters transcribe the *trp* operon of the supraoperon, one preceding the first gene of the supraoperon, *aroF*, and one preceding the *trp* operon region. The seventh *trp* gene of *B. subtilis*, *trpG*, is not within the *trp* operon; it resides in the folate operon, where it functions in both tryptophan and folate formation [2]. Transcription of the *trp* operon region of the *B. subtilis* supraoperon is regulated by a mechanism of transcription attenuation that is based on the ability to form two RNA-secondary structures: the alternative antiterminator and the transcription terminator [3,4]. A tryptophan-activated RNA-binding protein, TRAP, chooses between these alternative structures. In its tryptophan-activated state, TRAP binds to (G/U)AG RNA-triplet repeats, some of which are located within the segment of the transcript and specify the antiterminator. TRAP binding disrupts the antiterminator, enabling the transcription terminator to form and terminate transcription. [5]. Thus, whenever tryptophan is growth limiting, TRAP is inactive, and the antiterminator structure persists and transcription continues so that the structural genes of the operon are transcribed [6]. Uncharged $tRNA^{Trp}$ is also sensed as a regulatory signal in *B. subtilis*; its accumulation leads to increased *trp* operon expression [7,8,19]. This is due to the action of an additional regulatory protein, Anti-TRAP (AT). AT can inhibit the ability of TRAP to function [8–10]. Synthesis of AT, a product of the separate *at* operon, is based on tandem transcriptional and translational sensing of uncharged $tRNA^{Trp}$ [10]. Transcription of the *at* operon is regulated by the T-box transcription-attenuation mechanism, which is based on uncharged $tRNA^{Trp}$ binding to – and stabilizing – a transcription-antiterminator structure [10–12]. Thus, transcription of the *B. subtilis* *trp* operon is based on mechanisms that sense tryptophan and uncharged $tRNA^{Trp}$.

Analysis of *trp*-operon regulation in Gram-positive bacteria

To what extent is *trp*-operon organization conserved in Gram-positive species? Are L-Trp and uncharged-$tRNA^{Trp}$ conserved as signal molecules in regulating transcription of the biosynthetic *trp* operons of all Gram-positive species? What regulatory mechanisms are employed; and are they identical? To answer these questions, we first searched all the complete Gram-positive bacterial genome sequences available in the GenBank database (ftp://ftp.ncbi.nih.gov/GenBank/) for the existence of TRAP and AT orthologs. We found TRAP homologs in the genomes of *B. subtilis, Bacillus clausii*, *Bacillus halodurans, Bacillus licheniformis, Bacillus pumilus, Clostridium thermocellum, Bacillus stearothermophilus, Bacillus kaustophilus, Moorella thermoacetica* and *Oceanobacillus iheyensis*. An AT-encoding gene was observed only in *B. subtilis* and *B. licheniformis* (Figure 1). Although the existence of a TRAP protein was predicted in *O. iheyensis*, there were no discernible TRAP-binding-site triplets in the transcript of

Corresponding author: Merino, E. (merino@ibt.unam.mx).
Available online 13 June 2005

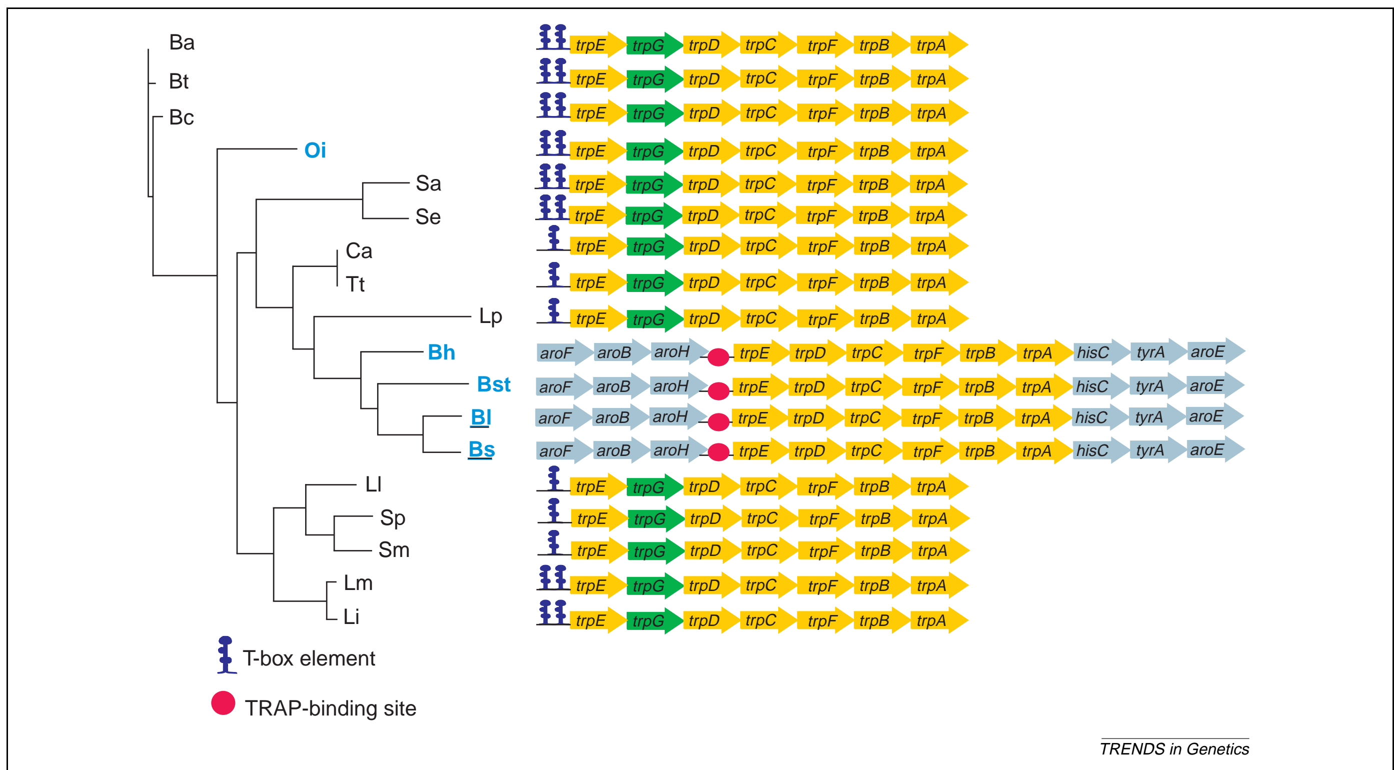

Figure 1. The organization of tryptophan biosynthetic operons in different Gram-positive bacteria; the relationships of their phylogeny, genome context and transcription regulatory elements are shown. To detect possible evolutionary and regulatory relationships among the *trp* biosynthetic operons of different fully-sequenced Gram-positive genomes, we used the MrBayes program [21] to construct a tree based on the concatenation of the TrpE-TrpD-TrpC-TrpF-TrpB-TrpA protein sequences of each genome. The regulatory elements were identified using the MAST program [22] and the conserved motifs defined in Ref. [23]. The genome context of the *trp* operons of different organisms was analyzed using our web server GeConT [24]. The names of organisms in blue have a gene encoding TRAP ortholog in their genomes. The names of organisms that are underlined correspond to genomes with anti-TRAP (AT) orthologs. A red circle identifies the location of a TRAP-binding site in the corresponding transcript. The *trpG* gene, present in all the tryptophan biosynthetic operons shown except those regulated by TRAP, is in green. In these TRAP regulated operons, *trpG* specifies a bifunctional protein and it resides within the folate operon. *trp* operons that are regulated by one T-box element are denoted with a single blue stem-loop structure, whereas *trp* operons regulated by tandem T-box elements are denoted by two blue stem-loop structures. The full name of the strains examined in these analyses are as follows: Ba, *Bacillus anthracis Ames*; Bc, *Bacillus cereus ATCC 10987*; Bh, *Bacillus halodurans*; Bl, *Bacillus licheniformis*; Bst, *Bacillus stearothermophilus*; Bs, *Bacillus subtilis;* Bt, *Bacillus thuringiensis konkukian*; Ca, *Clostridium acetobutylicum*; Li, *Listeria innocua*; Ll, *Lactococcus lactis*; Lm, *Listeria monocytogenes*; Lp, *Lactobacillus plantarum*; Oi, *Oceanobacillus iheyensis*; Sa, *Staphylococcus aureus Mu50*; Se, *Staphylococcus epidermidis ATCC 12228*; Sm, *Streptococcus mutans*; Sp, *Streptococcus pneumoniae R6*; Tt, *Thermoanaerobacter tengcongensis*.

the leader region of the *trpEGDCFBA* operon or elsewhere in its entire genome. Instead, transcription of the *trp* operon in *O. iheyensis* appears to be regulated by the T-box mechanism, a common tRNA-responsive attenuation mechanism that is used in used in regulating the expression of many tRNA synthetase and other operons in *B. subtilis* and in other Gram-positive bacteria [12,13]. Previous studies had detected T-box leader regions as potential *trp*-operon regulators in six Gram-positive bacterial species. In two of these, *B. anthracis and L. monocytogenes*, two sets of T-boxes were predicted to be arranged in tandem upstream of the *trpE* gene [14] (Figure 1). T-box elements contain several essential features: a stem-and-loop structure with a UGG tryptophan codon in a 'specifier loop', an antiterminator structure with a T-box consensus sequence that could pair with the CCA $tRNA^{Trp}$ acceptor arm and a Rho-independent transcription-terminator structure (Figure 2b) (reviewed in Ref. [12]). Our computer search of completely sequenced bacterial genomes extended this list of presumptive T-box-regulated *trp* operons to include 12 Gram-positive genomes (Figure 2a). In these organisms, there is as yet no evidence of a TRAP-like protein and there are no detectable TRAP-like RNA-binding sites. Do these organisms lack the dual sensing systems that are present in *B. subtilis?* Unexpectedly, we observed that in most of these presumed T-box-regulated operons (8/14), there is a second, adjacent T-box element in the 5′ segment of the leader transcript (Figure 1). In the 150 aminoacyl-tRNA synthetase operons that are regulated by the T-box mechanism in 30 fully-sequenced Gram-positive genomes, we observed only 11 operons that had T-boxes arranged in tandem.

All of the T-box elements identified in our study match the features proposed by Chopin [15] and Grundy and Henkin [12,13]. These features are: (i) T-box elements are present in a non-coding region; (ii) a typical intrinsic transcription terminator is present downstream of the T-box sequence (13–36 bases on the 3′ side of the T-box sequence); (iii) the RNA segment bearing the T-box sequence has the potential to fold into a consensus antiterminator structure; (iv) in this antiterminator, there is a loop called the 'antiterminator bulge' with a conserved UGG triplet that can pair with the CCA acceptor arm of the tRNA; (v) a start codon is present downstream of the T-box sequence (70–148 nucleotides downstream); (vi) a large stem-loop structure called 'Stem I' contains a 'specifier loop' with the corresponding

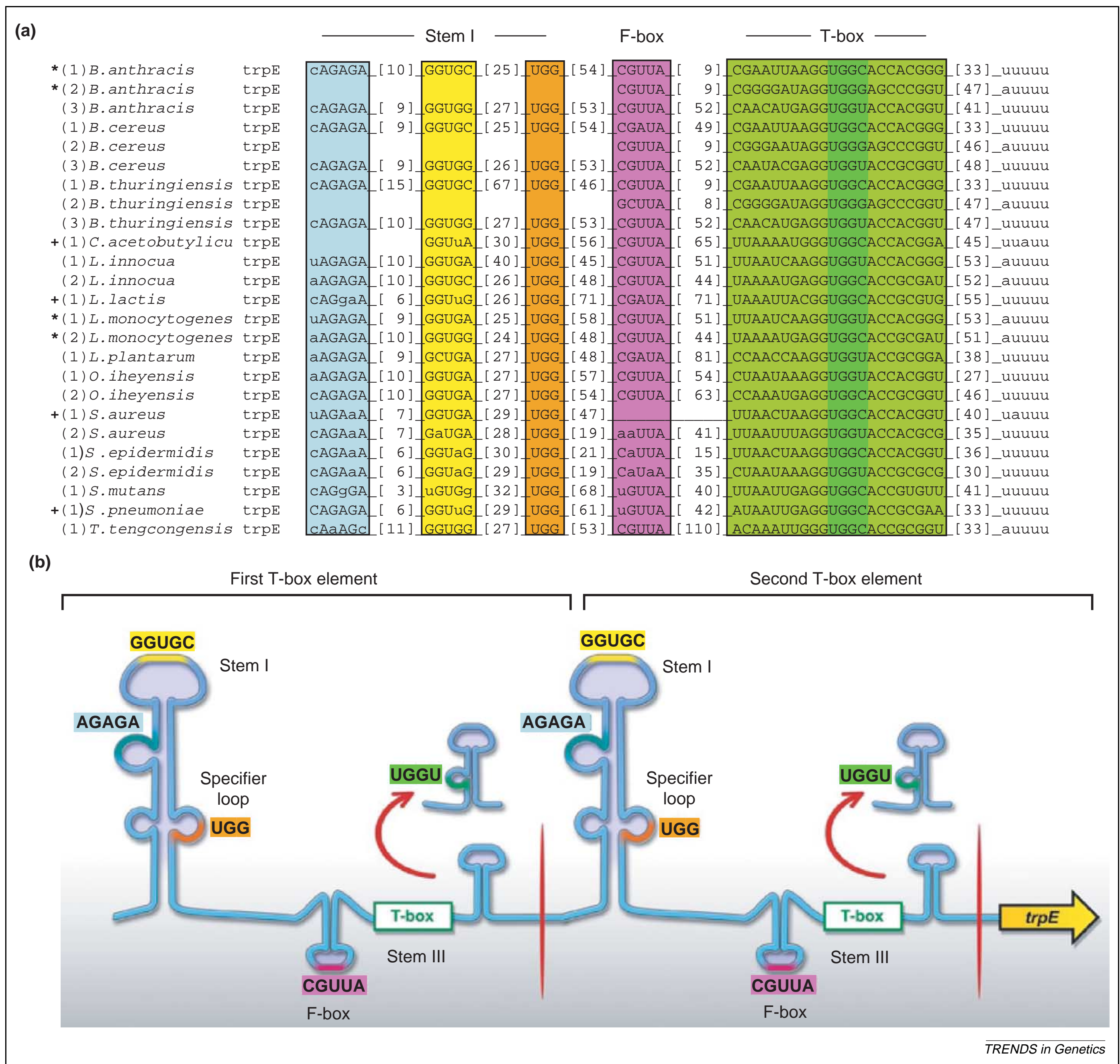

Figure 2. Predicted tandem T-box elements in the leader transcripts of *trp* biosynthetic operons of Gram-positive bacteria. **(a)** Sequence alignment of the leader transcripts of the *trp* biosynthetic operons identified in our analyses (Figure 1). Conserved motifs corresponding to previously reported important elements in the RNA sequences [12,13] are in boxes. Capital letters indicate conserved nucleotides. Organisms marked with an asterisk were previously reported to have a tandem T-box [14]. Organisms marked with a plus (+) were reported to have a *trp* operon that is regulated by a T-box [14,15,25]. The number in parenthesis indicates whether this is the first, second or third presumptive T-box, in the leader region of the operon. Note that the second T-box sequence cannot fold to form an antiterminator structure; therefore, it is not functional and might participate in an as yet undescribed regulatory mechanism (see main text). **(b)** Schematic representation of tandem T-box elements. Conserved elements in the structures are colored as in Figure 2a. A vertical red line indicates that if a terminator structure forms, transcription will be terminated. In the regulatory regions with three T-box sequences (*Bacillus anthracis, Bacillus thuringiensis* and *Bacillus cereus*), the second T-box sequence overlaps the specifier hairpin of the last T-box element.

codon that determines the specificity of the amino acid response (UGG for Trp, in our case); (vii) the conserved sequences AGAGA and G(G/C)UG(C/A) are appropriately placed in the loop regions of Stem I; (viii) a conserved stem-and-loop structure with the F-box sequence (CGUUA) is essential, and experimental data indicate that mutations in this sequence result in a dramatic reduction in basal level activity [16]; and (ix) the GA motif (as defined by two short helices separated by an asymmetric internal loop, with highly conserved GA dinucleotide sequences positioned on opposite sides of the internal loop) is located more often in T-box sequences in the *Bacillus*/*Clostridium* group than in the *Streptococcus* or the *Staphylococcus* groups [17]. In the T-box-regulated *trp* operons, the GA motif is found only in members of the *Bacillus*/*Clostridium* group.

In the *trp* operons of *B. anthracis*, *B. cereus* and *B. thuringiensis,* three T-box-like sequences are present. The first and the third meet all of the criteria mentioned previously. The second T-box in these operons might not be

a functional T-box, or it might participate in an as yet undescribed regulatory manner, because it cannot fold to form an antiterminator structure, its upstream sequence could not fold into a specifier hairpin structure and it lacks other T-box features (Figure 2).

The existence of different regulatory strategies for the same gene or operon is likely to be a consequence of different selective pressures evoked by individualistic metabolic inter-relationships that distinguish various organisms. Interestingly, in our study, we observed that the distinct regulatory elements and mechanisms proposed for the regulation of *trp*-operon transcription correlate with variations in operon organization. Thus, *trp* operons regulated by the T-box mechanism have the *trpEGDCFBA* operon organization, whereas *trp* operons regulated by TRAP have lost *trpG* from the genome, and the remaining *trp*-operon region appears to have been inserted into an *aro* operon. This produces the *aro* supra-operon, *aroFBH-trpEDCFBA-hisC-tyrA-aroE* (Figure 1) [18]. These differences in operon organization that can relate the specific *trp* pathway with a larger metabolic capability might well have been responsible for the development of different regulatory strategies. Curiously, *Oceanobacillus iheyensis* has both a T-box-regulated *trp* operon and TRAP (but no obvious TRAP targets). The organization of the *O. iheyensis trp* operon conforms to the ancestral organization of low-GC Gram-positive bacteria proposed by Xie *et al.* [18]. At least three evolutionary possibilities can be offered at present. (i) TRAP might have been acquired as a 'stray' protein and might be a result of recent lateral gene transfer. This would be consistent with the absence of TRAP-binding sites. (ii) TRAP might have been originally acquired in an ancestral organism similar to certain *Listeria* species because it has a tandem T-box-regulated *trp* operon. It might represent an intermediate between *Listeria*, for example, and the *Bacillus* clade that possesses functional TRAP-binding sites. (iii) TRAP in *O. iheyensis* might be a remnant of an ancestral state where competence for both TRAP and T-box regulation co-existed, with one being subsequently lost in different lineages. One can anticipate that as a greater genome representation becomes available in the near future, phylogenetic gaps among organisms will be filled and the most probable scenario will be revealed.

Potential regulatory role of tandem *trp* T-box elements

Transcription of the structural gene region of operons of many species is regulated by two or more sequential events. An advantage of using tandem regulatory decisions is that it would expand the range of expression of the respective operon. In our study, we observed that the *trp* operon of several bacterial species contains tandem *trp* T-box elements in its leader transcript segment preceding *trpE*. If these organisms recognize tryptophan as a regulatory signal, it must be by a mechanism other than TRAP function. However, the presence of tandem *trp* T-box elements in the *trp* operon leader region of certain species would enable the *trp* operon to have a significantly greater range of transcriptional regulation in response to changes in the level of uncharged $tRNA^{Trp}$. A similar example of the possible use of tandem regulatory elements has recently been described by Mandal *et al.* [20]. Obviously, experimental analyses must be performed with each of these systems to assess these possibilities.

Concluding remarks and perspectives

Our analyses reveal considerable flexibility in the design of mechanisms used for regulating transcription of the tryptophan biosynthetic operons of Gram-positive bacteria. The organisms included in our study fall into three classes: (i) species with a *trp* operon regulated by TRAP and AT (AT is regulated by a T-box). Only *B. subtilis* and *B. licheniformis* are in this class; (ii) species with a *trp* operon regulated by TRAP only; and (iii) species with a *trp* operon that is not regulated by TRAP but is regulated by one or more T-box elements. This third category was the one most commonly found. Whether both T-boxes in the leader RNA of the *trp* operons of some species are functional, and whether their presence permits a greater range of regulation of *trp* operon expression, as we propose, remains to be established.

Acknowledgements

We thank Randy Berka for providing the genome sequence of *B. licheniformis* before its publication. We also thank Shirley Ainsworth for bibliographical assistance, Guillermo Preciado for helping with the art work, and Lorenzo Segovia for helping with the phylogenetic analysis. We also thank Abel Linares, Arturo Ocadiz, Juan Manuel Hurtado and Alma Martinez for computer support. This study was supported by CONACyT grant 44213-Q to E.M., NSF Grant MCB-0093023 to C.Y. and NLM grant LM008297 to R.J.

References

1 Yanofsky, C. (2004) The different roles of tryptophan transfer RNA in regulating *trp* operon expression in *E. coli* versus *B. subtilis*. *Trends Genet.* 20, 367–374
2 Gollnick, P. *et al*. (2002) Aromatic amino acid metabolism in *Bacillus subtilis*. In *Bacillus subtilis and its Closest Relatives: From Genes to Cells* (Sonenshine, A.L. *et al*., eds), pp. 233–244, ASM Press
3 Henner, D.J. and Yanofsky, C. (1993) Biosynthesis of aromatic amino acids. In *Bacillus subtilis and Other Gram Positive Bacteria: Biochemistry, Physiology and Molecular Genetics* (Losick, R., ed.), pp. 269–280, ASM Press
4 Shimotsu, H. *et al*. (1986) Novel form of transcription attenuation regulates expression the *Bacillus subtilis* tryptophan operon. *J. Bacteriol.* 166, 461–471
5 Babitzke, P. *et al*. (1994) TRAP, the *trp* RNA-binding attenuation protein of *Bacillus subtilis*, is a multisubunit complex that appears to recognize G/UAG repeats in the *trpEDCFBA* and *trpG* transcripts. *J. Biol. Chem.* 269, 16597–16604
6 Babitzke, P. and Yanofsky, C. (1993) Reconstitution of *Bacillus subtilis trp* attenuation *in vitro* with TRAP, the *trp* RNA-binding attenuation protein. *Proc. Natl. Acad. Sci. U. S. A.* 90, 133–137
7 Steinberg, W. (1974) Temperature-induced derepression of tryptophan biosynthesis in a tryptophanyl-transfer ribonucleic acid synthetase mutant of *Bacillus subtilis*. *J. Bacteriol.* 117, 1023–1034
8 Lee, A.I. *et al*. (1996) A temperature-sensitive *trpS* mutation interferes with *trp* RNA-binding attenuation protein (TRAP) regulation of *trp* gene expression in *Bacillus subtilis*. *J. Bacteriol.* 178, 6518–6524
9 Valbuzzi, A. and Yanofsky, C. (2001) Inhibition of the *B. subtilis* regulatory protein TRAP by the TRAP-inhibitory protein, AT. *Science* 293, 2057–2059
10 Chen, G. and Yanofsky, C. (2003) Tandem transcription and translation regulatory sensing of uncharged tryptophan tRNA. *Science* 301, 211–213
11 Grundy, F.J. and Henkin, T.M. (2002) Synthesis of serine, glycine,

cysteine and methionine. In *Bacillus subtilis and its Closest Relatives: From Genes to Cells* (Sonenshine, A.L. *et al*., eds), pp. 245–254, ASM Press
12 Grundy, F.J. and Henkin, T.M. (2003) The T box and S box transcription termination control systems. *Front. Biosci.* 8, D20–D31
13 Grundy, F.J. *et al*. (2002) Sequence requirements for terminators and antiterminators in the T box transcription antitermination system: disparity between conservation and functional requirements. *Nucleic Acids Res.* 30, 1646–1655
14 Panina, E.M. *et al*. (2003) Regulation of biosynthesis and transport of aromatic amino acids in low-GC Gram-positive bacteria. *FEMS Microbiol. Lett.* 222, 211–220
15 Chopin, A. *et al*. (1998) Analysis of the *Bacillus subtilis* genome sequence reveals nine new T-box leaders. *Mol. Microbiol.* 29, 662–664
16 Rollins, S.M. *et al*. (1997) Analysis of *cis*-acting sequence and structural elements required for antitermination of the *Bacillus subtilis tyrS* gene. *Mol. Microbiol.* 25, 411–421
17 Winkler, W.C. *et al*. (2001) The GA motif: an RNA element common to bacterial antitermination systems, rRNA and eukaryotic RNAs. *RNA* 7, 1165–1172
18 Xie, G. *et al*. (2003) Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol. Mol. Biol. Rev.* 67, 303–342
19 Sarsero, J.P. *et al*. (2000) A *Bacillus subtilis* operon containing genes of unknown function senses $tRNA^{Trp}$ charging and regulates expression of the genes of tryptophan biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 97, 2656–2661
20 Mandal, M. *et al*. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306, 275–279
21 Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574
22 Bailey, T.L. and Gribskov, M. (1998) Combining evidence using *P*-values: application to sequence homology searches. *Bioinformatics* 14, 48–54
23 Abreu-Goodger, C. *et al*. (2004) Conserved regulatory motifs in bacteria: riboswitches and beyond. *Trends Genet.* 20, 475–479
24 Ciria, R. *et al*. (2004) GeConT: gene context analysis. *Bioinformatics* 20, 2307–2308
25 van de Guchte, M. *et al*. (1998) tRNA(Trp) as a key element of antitermination in the *Lactococcus lactis trp* operon. *Mol. Microbiol.* 29, 61–74

0168-9525/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tig.2005.06.001

Striking nucleotide frequency pattern at the borders of highly conserved vertebrate non-coding sequences

Klaudia Walter[1,2], Irina Abnizova[1,2], Greg Elgar[1] and Walter R. Gilks[2]

[1]MRC Rosalind Franklin Centre for Genomics Research, Hinxton, Cambridge, UK, CB10 1SB
[2]MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge, UK, CB2 2SR

In a recent study, 1373 highly conserved non-coding elements (CNEs) were detected by aligning the human and *Takifugu rubripes* (Fugu) genomes. The remarkable degree of sequence conservation in CNEs compared with their surroundings suggested comparing the base composition within CNEs with their 5′ and 3′ flanking regions. The analysis reveals a novel, sharp and distinct signal of nucleotide frequency bias precisely at the border between CNEs and flanking regions.

Introduction

An ever-growing number of whole genome sequencing projects are being completed, enabling genome comparisons through a variety of different bioinformatics tools [1]. The human genome comprises ~3000 Mb of DNA, the greatest part of which does not encode proteins (>98%) [2], whereas the vertebrate genome of *Takifugu rubripes* (Fugu, Japanese pufferfish) contains only 400 Mb. Nevertheless, Fugu and human, whose last common ancestor lived 450 million years ago, have a similar gene repertoire [3–5]. This makes Fugu a particularly interesting genome for comparison with human, especially with regard to the identification and characterisation of cis-regulatory regions such as enhancers or silencers within non-coding DNA.

Recently, 1373 highly conserved non-coding elements (CNEs) were discovered by aligning the human and Fugu genomes [6]. These CNEs show a remarkable degree of sequence conservation with an average identity of 84.3% over lengths ranging 93–740 bp (an average of 199 bp). CNEs are found in all human chromosomes, except for chromosomes 21 and Y. They generally appear in clusters (e.g. 85% of CNEs are within 370 kb of the next CNE) [6]. More than 93% of the clusters are located within 500 kb of genes involved in transcriptional regulation or development, often located in regions of low gene density [6]. Most CNEs are also conserved in the mouse, rat, chicken and zebrafish genomes [6].

Nucleotide sequence patterns are characteristic of several genome punctuation marks, for instance, sharp changes in GC-content at transcription boundaries [7] or the signal at intron donor and acceptor splice sites. The variation in nucleotide frequency patterns across genes has been comprehensively illustrated in Louie *et al.* [8]. However, CNEs are distinct from coding regions and other expressed sequences. We have analysed the base composition of human and Fugu CNEs along with their flanking regions to identify any characteristic patterns or signatures. We have compared these with patterns detected around the coding regions of genes.

Corresponding author: Gilks, W.R. (wally.gilks@mrc-bsu.cam.ac.uk).
Available online 23 June 2005