

Erratum

Transcription attenuation in bacteria: theme and variations

Magali Naville and Daniel Gautheret

Brief Funct Genomic Proteomic 2009;8:482–492

In the above article, two references were removed in error and subsequent citations numbered incorrectly. The below reprint of this article reinstates the references and corrects the citations. The publisher apologises for this error.

Transcription attenuation in bacteria: theme and variations

Magali Naville and Daniel Gautheret

Advance Access publication date 3 August 2009

Abstract

Premature termination of transcription, or attenuation, is an efficient RNA-based regulatory strategy that is commonly used in bacterial organisms. Attenuators are generally located in the 5' untranslated regions of genes or operons and combine a Rho-independent terminator, controlling transcription, with an RNA element that senses specific environmental signals. A striking diversity of sensing elements enable regulation of gene expression in response to multiple environmental conditions, including temperature changes, availability of small metabolites (such as ions, amino acids, nucleobases or vitamins), or availability of macromolecules such as tRNAs and regulatory proteins. The wide distribution of attenuators suggests an early emergence among bacteria. However, attenuators also display a great mobility and lability, illustrated by a multiplicity of recent horizontal transfers and duplications. For these reasons, attenuation systems are of high interest both from a fundamental evolutionary perspective and for possible biotechnological applications.

Keywords: *Attenuation; bacteria; transcriptional termination; gene regulation; evolution*

Corresponding author. Daniel Gautheret, Institut de Génétique et de Microbiologie, Paris-Sud University, Bâtiment 400, F-91405 Orsay Cedex, France. Tel: +33-1-69-15-46-32; Fax: +33-1-69-15-46-29; E-mail: daniel.gautheret@u-psud.fr

Magali Naville is a PhD candidate at the Institut de Génétique et de Microbiologie, Paris-Sud University, and is working on computational prediction and analysis of small regulatory RNAs in Prokaryotes.

Daniel Gautheret is a Professor of Molecular Biology at the Institut de Génétique et de Microbiologie, Paris-sud University. His research focuses on the sequence, structure and function of non-coding RNA.

INTRODUCTION

Living organisms have developed gene expression control strategies that interfere with virtually every step of the transcription and translation processes as well as with mRNA degradation, protein stability and localization or protein–protein interactions. In bacteria, a crucial control point of gene expression occurs just after transcription is initiated, as the RNA polymerase complex progresses through the 5′ untranslated region (UTR) of a gene. A large number of transcription units contain premature ‘terminator’ structures that conditionally interrupt transcription before a functional protein-coding RNA is formed. This mechanism was first discovered in the tryptophan operon [1] where the 5′ terminator is associated to a leader sequence that renders it responsive to the cellular availability of tryptophan, through the mediation of ribosomes and Trp-tRNAs. The complete leader-terminator structure was called an ‘attenuation’ system and was later discovered to also apply to other amino acid synthesis operons. Since this initial discovery, 5′ terminators have been found to be associated with a wide range of sensing mechanisms and biological functions. In this review, we will follow Yanofsky [2] in classifying all such regulatory systems under the term ‘attenuators’.

Sensu largo, we define attenuators as 5′-*cis* regulatory elements that can fold into two alternative structures, one of them consisting of a Rho-independent terminator (Figure 1). Attenuation then involves the correct folding of the RNA terminator structure at a site located between the promoter and the first gene of the operon. Around this simple scheme, bacteria developed a markedly wide range of systems, each of them coupling formation of the termination structure to some sensing mechanism able to detect a determining environmental factor or event. The exquisite diversity of sensing mechanisms may not have been possible without the remarkable structural abilities of the RNA molecule, which is particularly suited to signal detection, since it both enables highly specific molecular recognition and can undergo conformational changes, ensuring transduction of the signal. The availability of hundreds of complete bacterial genomes now provides an opportunity to reexamine the expanding attenuator collection from a comprehensive evolutionary perspective.

Rho-independent, or intrinsic, terminators are used primarily at the 3′ end of genes and represent highly inexpensive termination mechanisms, since

they do not require any additional protein factor to mediate transcription termination. At the DNA level, Rho-independent terminators generally consist of a GC-rich palindrome followed by a run of T’s defining the site of termination. At the RNA level, the palindrome gives rise to a stable stem-loop structure followed by a run of U’s. While the hairpin induces pausing of the transcribing RNA polymerase complex, the relatively weak binding between the poly U tail in the nascent RNA transcript and the corresponding poly A sequence in the DNA presumably causes dissociation of the polymerase from the DNA template, releasing the nascent mRNA and terminating transcription. The poly U tail may also induce polymerase stalling, thereby providing time for the RNA hairpin to form [3]. These characteristics facilitate the computational detection of intrinsic termination signals, albeit not in all organisms. For example, in the Gram-positive species *Streptomyces lividans*, the few experimentally identified terminators lack a run of U’s following the stem-loop [4]; this observation could also apply to other species [5].

VARIATIONS ON A THEME

We propose in this review to present attenuation systems by order of increasing complexity of the associated sensing mechanism.

Attenuation-based RNA thermometers

These 5′-UTR elements control expression of downstream genes in a temperature-dependent manner (see [6] for review). Heat-shock, cold-shock and virulence genes constitute typical targets for RNA thermometers. In pathogenic organisms they allow an induction of virulence genes expression in response to the entry of the bacterium into its host. Present knowledge relates RNA thermometers to translation regulation rather than transcriptional attenuation. Indeed, all such elements known in 2007 act by controlling the ribosome accessibility of a Shine-Dalgarno sequence [6, 7], a well-studied case being the ROSE (Repression Of heat Shock gene Expression) element [8]. However, our domain wide prediction, presented below, indicates that intrinsic terminators precede many genes of the Cold Shock Protein E (CspE) family. This observation supports recent analysis demonstrating cold-shock induction of *cspE* in *Escherichia coli* [9], which is mediated by formation of alternative secondary

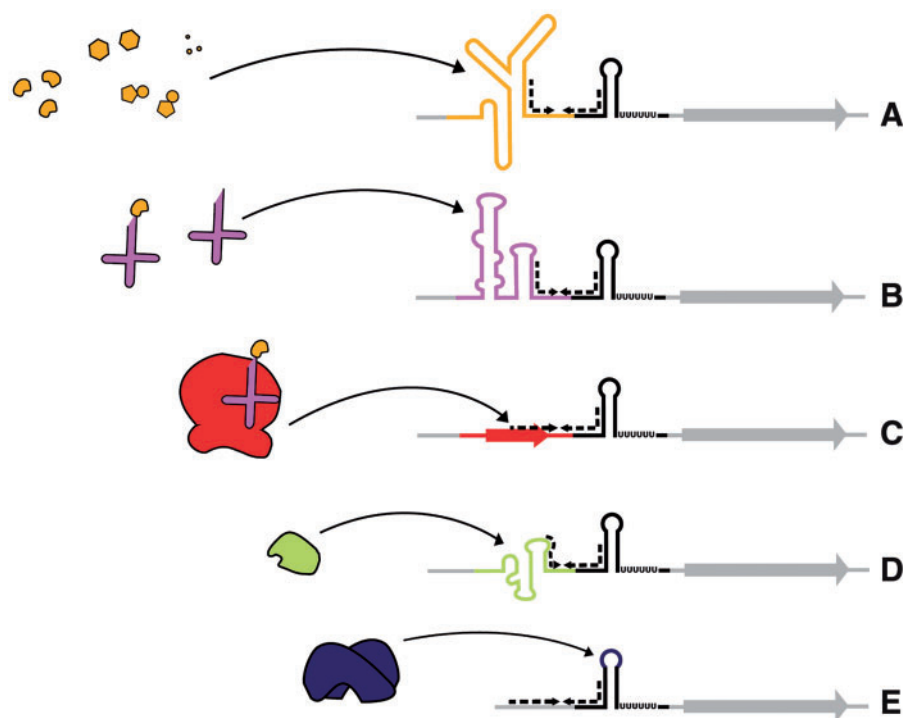


Figure 1: The major classes of bacterial attenuators and their mechanisms. **(A)** Riboswitch. The RNA aptamer binds directly a specific small ligand (different ligand types are symbolized to the left: amino acids, sugars, nucleobases or ions). The ligand concentration is deciding for the riboswitch folding into a terminator (as shown) or an anti-terminator structure (dotted arrows). **(B)** T-box. The RNA element folds into an anti-terminator (dotted arrows) upon interaction with a specific uncharged tRNA (symbolized to the left). If charged tRNAs, unable to interact with the T-box, predominate, the terminator forms (as shown). **(C)** Peptide leader. The alternative folding into a terminator (as shown) or an anti-terminator (dotted arrows) structure depends on the translation rate of a short ORF enriched in particular codons (short arrow) by the ribosomal machinery (symbolized to the left). This translation rate depends in turn on the concentration of corresponding charged tRNAs available. **(D)** Ribosomal protein leader. The alternative folding relies on the interaction of the RNA leader with a particular ribosomal protein (symbolized to the left), if available in the cytosolic environment. The RNA aptamer mimics a ribosomal RNA interacting with the considered protein in the ribosome structure. **(E)** Attenuation by binding of a terminator/anti-terminator protein. The alternative terminator/anti-terminator folding is controlled by a specific interaction with an RNA-binding protein (symbolized to the left).

structures in the 5'-UTR of *cspE* transcript at 37°C and 15°C. CSP's have been shown to exhibit both DNA and RNA binding activity, intervening as transcription factor and RNA chaperones. Moreover, CspA, CspC and CspE exhibit a transcription anti-termination activity [10] that could be intriguingly linked to their own attenuation-regulating element.

Riboswitches that function by forming alternative termination/anti-termination structures (Figure 1A)

Riboswitch sequences sense the level of a particular small metabolite in the cell through direct ligand–

RNA interaction. Upon binding of the ligand to a domain called aptamer, a conformational change transduces the signal via an expression platform, which can be either a Rho-independent terminator or a hairpin blocking the ribosome-binding site. Several reviews have been devoted to these regulatory elements [11–14], that can detect a variety of metabolites: amino acids and derivatives (Lysine, Glycine, SAM), nucleosides and derivatives (Guanine, Adenine, PreQ1, cyclic di-GMP [15]), sugars (Gln6P), vitamin cofactors (AdoCbl, TPP, FMN, Moco [16]) and metal ions (Mg^{2+} [17]). The aptamer domain forms a complex 3D scaffold that is further structured by binding of the ligand. The length of the aptamer usually varies between

100 and 200 bp, with the exception of the 40-bp PreQ1 element. Most riboswitches have an inhibitory effect on genes encoding biosynthetic enzymes or transporters for example, of which synthesis should stop when corresponding metabolites reach sufficiently high concentrations. However, certain instances of glycine, adenine and lysine riboswitches achieve ‘ON activity’, through formation of a structure that interferes with the terminator upon metabolite binding. Previous studies have distinguished two types of control by these elements, either thermodynamic or kinetic, depending on the rate of transcription elongation by RNA polymerase on one hand and on the free energy of the structure adopted by the transcript, on the other hand (see [12] for review).

Functional knowledge on riboswitches has greatly benefited from crystal structures obtained in the past 4 years (see [13] for a complete list). Among the most recent structures are lysine riboswitches [18, 19], characterized by a long and complex aptamer. The precise description of this riboswitch’s scaffold and its conformational changes upon amino acid and analogs binding appears promising for the development of appropriate inhibitory pharmaceuticals. Riboswitches generally contribute to cell homeostasis by maintaining physiological levels of sensed metabolites. Regulation by riboswitches is a world in itself, in which considerable diversity of combinations, such as tandem arrangements, is observed [20]. In the particular case of purine recognition by riboswitches, it is interesting to point out that the specificity for guanine or adenine binding is ensured by a single nucleotide (C or U, respectively) in the aptamer core, while the rest of the aptamer remains highly conserved in both sequence and structure [21]. Recently, the range of sensed metabolites has been further increased with the uncovering of riboswitches responding to the second messenger di-GMP [15] and to the molybdenum cofactor [16]. This list is rapidly expanding with the number of sequenced genomes and the advent of generalized transcriptomics.

T-boxes (Figure 1B)

These elements correspond to *cis*-acting regulators that undergo a structural shift in response to the binding of a specific uncharged tRNA (see [22] for a detailed review, and [23] for a complete view of networks controlled by T-boxes). While some authors consider T-boxes as another family of

riboswitches [22], we prefer to distinguish them here, since the RNA–RNA interaction that they rely on should be considered as a step beyond ‘small molecule’ riboswitches, in complexity. Anti-termination occurs when a sufficient level of a particular uncharged tRNA is achieved. The specific pairing of this tRNA is ensured by an “anti-anti-codon”, or specifier, presented by a particular loop, and promotes the stabilization of the anti-terminator structure, leading to expression of downstream genes. This system, mostly found in Firmicutes, often controls genes encoding aminoacyl tRNA synthetases as well as genes specifying proteins involved in amino acid biosynthesis and transport [23]. Their high sequence and structure conservation allows a relatively easy identification [23] and helps determine the specificity of the flanking gene. Structure and kinetics of T-boxes are now precisely understood. As for riboswitches, several occurrences of adjacent T-boxes resulting from recent duplications are observed, point mutations in the specifier codon ensuring the specificity change.

Amino-acid operon leaders (Figure 1C), or attenuators *sensu stricto*

As previously described elements, these leader RNAs’ can assume two alternative terminator/anti-terminator structures, but here the sensor element is a short peptide coding sequence that is enriched in codons for a specific amino acid [24]. The terminator forms when the cell contains an excess of this amino acid and therefore of the charged cognate tRNAs, facilitating ribosomal readthrough during synthesis of the leader peptide, which in turn prevents anti-terminator formation. Alternatively, a deficiency of a specific charged tRNA causes ribosome stalling on the leader peptide coding region and subsequent formation of the terminator structure, leading to release of the transcribing RNA–polymerase molecule that just synthesized the leader peptide. This family of attenuators is of particular interest since they illustrate the possible interactions, in bacteria, between transcription and translation, allowed by the colocalization of these two processes. Previously this has led authors to describe such a system as an example of ‘ribosome-mediated’ transcription attenuation [2, 25] that indirectly senses aminoacyl-tRNA charging (in comparison to T-boxes, which sense them directly), with the ribosome serving as the mediator.

Transcription attenuation by RNA–protein interactions

Ribosomal protein leaders controlling transcription (Figure 1D)

While rRNA transcription initiation mainly depends on the nucleoside triphosphate concentrations, ribosomal proteins transcription regulation is directly linked to the cellular concentration of rRNA. In Enterobacteria, the regulatory mechanism involves a structural similarity between the rRNA molecule on one hand, and the leader of the gene specifying the ribosomal protein on the other hand. A high protein level leads to its binding not only to every available rRNA molecules with which it interacts in the ribosome, thus contributing to ribosome assembly, but also to the 5'-leader of its own transcripts, that mimics its cognate rRNA region [26, 27]. Most ribosomal operons encode such a bifunctional protein, acting both as ribosomal scaffolding element and as regulator of transcription of the operon. Although these leaders are mainly associated with translation initiation mechanisms, some instances of transcriptional attenuation are also observed. Another system coordinating protein and RNA production through a mimicry between two RNA molecules is the Thr-tRNA-synthetase leader region of *E. coli* that mimics a tRNA, which is the substrate of the enzyme [28].

Attenuation by terminator/anti-terminator binding proteins (Figure 1E)

This type of attenuation is modulated by protein–RNA interaction, affecting the folding of the terminator. The regulatory protein activity is in turn controlled by its concentration in the cell or its activation status. Many examples of such systems can be cited, including the CSPs described previously, expression of which is temperature-dependent, or the PyrR protein found in Gram-positive bacteria, which regulates pyrimidine nucleotide biosynthesis operons [29] in response to the available level of UTP. If UMP, the product of the pathway, builds up in the cell, the PyrR termination factor is activated and induces transcriptional attenuation, mediating feedback inhibition of this pathway. In the same manner, and as for many other regulatory elements of two-component systems (TCSs), the BglG factor acts as an anti-terminator, leading to enhanced transcription under the activation of the sensory protein BglF, a β -glucoside-specific permease [30].

Potentially, for every system described above, an equivalent can be formed using a hairpin blocking the ribosome-binding site (RBS) in place of the terminator. Such declination further broadens the field of regulation possibilities. Evolutionary exchanges between the two regulation types are observed in a number of riboswitches, thermosensors and ribosomal protein leaders. The fact that such exchanges are not observed for certain elements raises an interesting evolutionary question.

PREMATURE TERMINATORS: A PAN GENOME ASSESSMENT

Genome-wide searches of intrinsic terminators have revealed large variations in the utilization of terminator systems, first among bacterial phyla but also among species of a given genus. For example, it has been shown that, in the Firmicutes phylum, the majority of terminators are Rho-independent (>90% in *Bacillus subtilis*) [31], whereas in *E. coli*, a Proteobacteria, transcriptional termination is achieved in a comparable manner by Rho-independent and Rho-dependent terminators. A predominant use of intrinsic termination was also observed in Neisseria, Psychrobacter, Pasteurellaceae and in the *Desulfovibrio desulfuricans* species [32]. The types of terminators involved in premature termination reflect this variation. Moreover, the complex interplay of auxiliary elements in Rho-dependent termination could suggest that such termination systems would be less appropriate than the simpler mechanism of intrinsic termination for use in premature termination, even though it has been shown in a few cases, such as the *tna* operon of *E. coli* [33].

Several attempts have been made at detecting different classes of attenuation mechanisms involving peptide leaders [24, 25, 34], riboswitches [35] or T-boxes [23]. Most of these predictions were based on the detection of conserved sequences preceding genes associated with the metabolism of specific compounds. Other approaches have attempted to detect conserved *cis*-regulatory non-coding RNAs, including terminator systems, independently of the function of the downstream genes [36]. In this review, we update the list of putative attenuators by performing a systematic search for terminators located upstream of translation start codons, and we sort the candidates according to the functional classification of regulated genes (Figure 2 and Supplementary Data). This search indicates that at

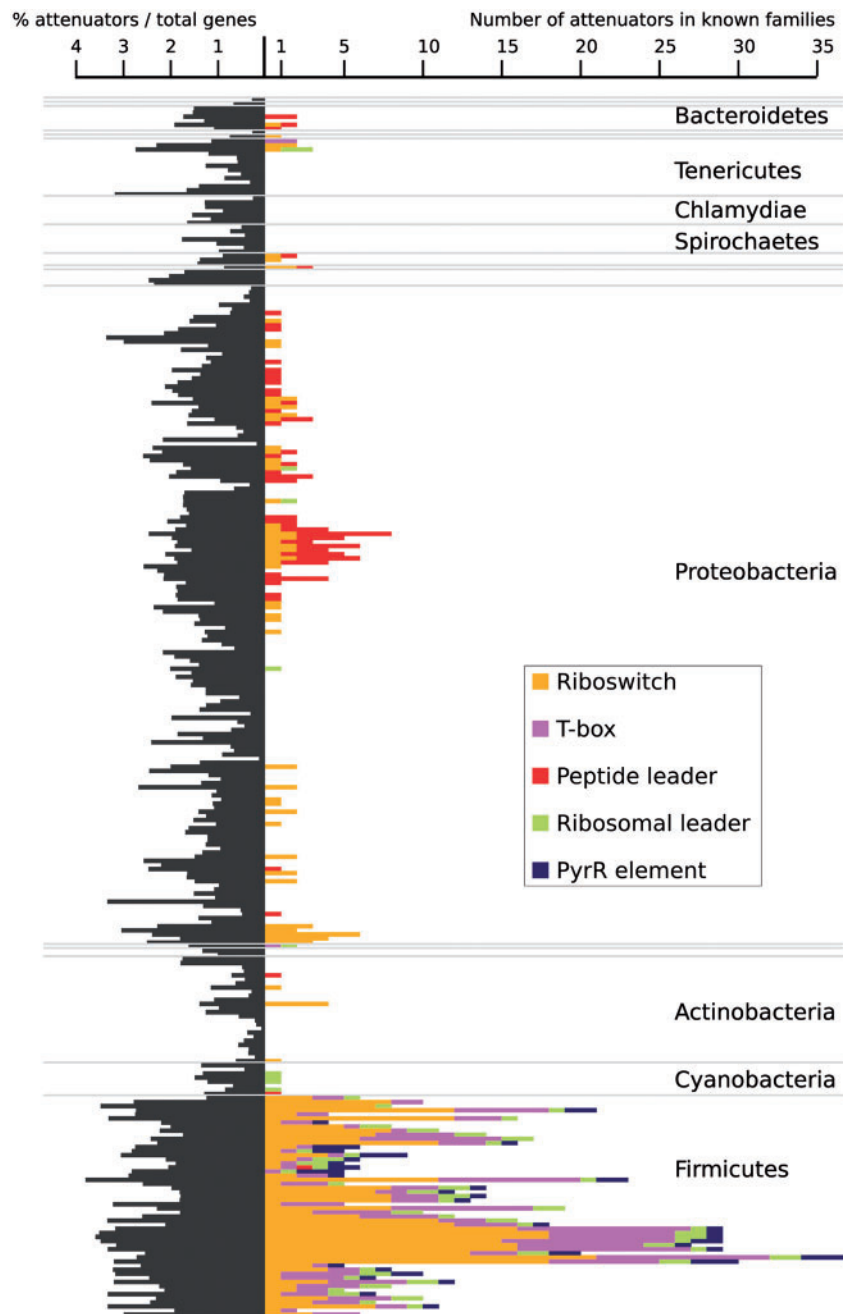


Figure 2: Distribution of attenuator classes across bacteria. Premature terminators were sought in the region 300 nt upstream of translation starts, using a combination of prediction programs (details of the procedures and search patterns are given in supplementary data). The left histogram shows the percentage of genes preceded by an attenuator in each genome. The right histogram shows absolute numbers of attenuators in each of the major attenuator classes.

least 1.6% of known bacterial genes are regulated by attenuation (Figure 2, left). This ratio ranges from 0.6% in Actinobacteria (*Streptomyces*) and 0.8% in Spirochaetes (*Treponema*) to 2.6% in Firmicutes (*B. subtilis*). The low ratio in Actinobacteria is probably linked to the high GC content of their genomes, which hampers motif search programs.

In Spirochaetes, however, the low premature terminator ratio may reflect the actual scarcity of these elements. In *E. coli*, 91 5'-terminators were identified, corresponding to 2.1% of its genes and ~3.3% of its operons regulated by attenuation. This number rises to 134 in *B. subtilis*, corresponding to 3.2% of its genes and about 5.0% of its operons. These

predictions may significantly underestimate the number of operons regulated by transcription attenuation, since it has been previously proposed that as many as 10% of the operons of certain species may be regulated by these mechanisms [37].

Further curation of the predicted attenuators using the Rfam database [38] and literature (Figure 2, right) indicates that the complete attenuator collection is far from elucidated. Only 4.9% of our predictions involve known elements, and these are unequally distributed: phyla such as Chlamydiae, Spirochaetes or Acidobacteria have almost no documented attenuators, whereas Proteobacteria have 2.1% and Firmicutes 16%. This discrepancy suggests that many attenuator functions remain to be identified.

FUNCTIONS CONTROLLED BY ATTENUATION

The great mechanistic diversity of transcription attenuators is a fertile ground for multiple usage by bacteria. A previous domain-wide analysis of attenuation systems [2] proposed a list of gene functions that are most frequently regulated by these bacterial mechanisms. Our own survey (Table 1) mostly

confirms this list, including as major targets of attenuation: the operon containing the transcription terminator–anti-terminator *musA*, the tryptophan biosynthetic operon, genes for aminoacyl-tRNA synthetases, genes regulated by riboswitches, ribosomes and RNA polymerase biogenesis operons as well as genes encoding transcriptional regulators of the LysR family. Pyrimidine metabolism is of particular interest regarding regulation by attenuation, since it is controlled by a wide range of terminator-based systems [40], illustrating the diversity of sensing possibilities. These include an UTP-sensitive peptide leader for *pyrBI* of *E. coli*, or attenuation by PyrR and reiterative initiation of transcription of *pyrG* [41] of *B. subtilis*. Particularly good ‘targets’ for attenuation regulation are also the translational and transcriptional factor genes *IF3* and *greA*. We distinguish three main functional categories of genes regulated by attenuation:

- (i) Regulation of nutrient biosynthesis and import: amino acids, nucleic acids, metal-ion homeostasis.
- (ii) Fluidity of housekeeping processes: regulation of tRNA-synthetase expression in function of

Table 1: Genes most frequently regulated by attenuation in bacteria

Hogonom family ^a	Score ^a	Regulated gene	Known attenuation systems ^b
HBG297802	131.6	yhbC (first gene of the nusA-infB operon, encoding an hypothetical protein)	nusA attenuator (<i>E. coli</i> ^c) (1/135)
HBG369991	72.3	rpLJ (50S ribosomal protein L10)	Ribosomal protein L10 leader (45/71)
HBG000748	66.9	pheS (phenylalanyl-tRNA synthetase)	T-box (23/59)/Peptide leader (10/59)
HBG181500	65.6	thrS (threonyl-tRNA synthetase)	T-box (21/69)/Peptide leader (14/69)
HBG142539	52.5	infC (translation initiation factor IF3)	Ribosomal protein L20 leader (31/48)
HBG046904	35.3	rpsL (30S ribosomal protein S12)	Ribosomal protein S12 leader
HBG343859	35.1	greA (transcription elongation factor GreA)	greA attenuator (<i>E. coli</i> ^c) (1/55)
HBG398889	31.0	hisS (histidyl-tRNA synthetase)	T-box (10/44)
HBG448785	29.6	trpE (anthranilate synthase)	T-box (3/34)/Peptide leader (21/34)
HBG249606	29.3	tyrS (tyrosyl-tRNA synthetase)	T-box (22/32)
HBG444146	29.1	metK (S adenosylmethionine synthetase)	SAM riboswitch (15/24)
HBG507301	28.3	pyrR (pyrimidine regulatory protein PyrR)	PyrR binding site (21/45)
HBG379584	28.2	ileS (isoleucyl-tRNA synthetase)	T-box (31/58)
HBG523338	28.1	pyrG (CTP synthase precursor)	Pyr regulatory elements
HBG114333	27.3	rpsP (30S ribosomal protein S16)	Ribosomal protein S16 leader
HBG001285	27.3	rpoB (DNA directed RNA polymerase beta chain)	rpL attenuator (<i>E. coli</i> ^c) (1/25)
HBG401792	26.9	cysE (serine acetyltransferase)	T-box (7/28)/SAM riboswitch (1/28)
HBG053760	26.5	serS (seryl-tRNA synthetase)	T-box (14/22)
HBG262099	25.9	leuA (2-isopropylmalate synthase)	Peptide leader (1/20)
HBG258034	25.2	pyrB (aspartate carbamoyltransferase)	PyrR-binding site (5/39)

^aFor each Hogonom gene family [39] and species, the number of detected attenuators was normalized by the total number of genes in this family in the considered species. Final scores correspond to the sum of normalized scores from each species. Families are sorted by decreasing scores.

^bNumbers of known/predicted attenuators are indicated in parentheses. Predicted attenuators are considered as ‘known’ if they match an entry of the Rfam database by a similarity search as described in the supplemental data.

^cAttenuation systems found in the literature to be described at least in *E. coli*. The number ‘1’ of known attenuators indicated in parentheses refers to the *E. coli* instance, but the exact number of described elements was not further evaluated.

available tRNAs; regulation of ribosomal operons and of the rRNA/ribosomal proteins balance for ribosome assembly.

- (iii) Global stress response: thermosensors, TCS sensitive to particular environmental signals.

A ROBUST AND INEXPENSIVE MECHANISM

From a mechanistic point of view, the common trait of attenuators is their ability to provide a fast and sensitive regulation of specific operons. However, a distinction should be made between systems that perform both sensing and expression regulation at the mRNA level and systems that outsource certain activities to a protein factor. In most attenuators, the signal sensing machinery that can bind the specific metabolite/code for the leader/fold differently according to the temperature is directly adjacent to the so-called expression platform (transcription terminator in this review, but also RBS sequestration in a broader sense). On the contrary, when a protein factor is involved, a decoupling occurs between the two players participating in the regulation. Constitutive expression of the sensor protein is necessary, since it must be constantly available for detection of physiological signals and induction of the appropriate genetic response. The development of proteins utilization by such regulatory systems, by the way it decouples the signal detection from the expression control, limits the direct detection by mRNA during elongation, and thus a detection at the close proximity of DNA. Therefore, in the way it increases DNA protection, regulatory protein utilization could contribute to maintain the integrity of genetic information. Furthermore, protein factors allow detection of extracellular events or components, thus enabling an anticipated answer, whereas RNA switches must 'wait' for the signal to reach the transcription machinery. This is one of the advantages of protein-based regulation, even if it is costly for the cell, as most protein sensors have to be continually turned over.

Interestingly, attenuators often act in combination with other types of regulation, and in particular with more global transcription factors. The use of dual mechanisms provides multiple spatio-temporal regulatory levels for any transcription unit, mediated by elements presenting different scales of reactivity and providing different speeds of response. One illustration of such concerted regulation is given by the

B. subtilis *xpt-pbuX* and *pbuG* genes, which are under dual regulation by a purine riboswitch and by the transcription factor PurR [11]. Likewise, the *alaRT* operon of the same species is regulated cooperatively by an Ala-T-box and by the transcriptional regulator AlaR [23].

Different types of attenuators may also cooperate efficiently in the regulation of the genes of a given metabolic pathway. An excellent example of such concerted action is given by the recently elucidated control of the genes of ethanolamine utilization in the Firmicute *Enterococcus faecalis* [42], involving two attenuation systems:

- terminators targeted by the response regulator EutV of the EutW-EutV TCS, that mediates anti-termination and allows sensing of extracellular ethanolamine;
- the AdoCbl riboswitch (containing an intrinsic terminator) that senses AdoCbl, a cofactor in the catabolism of ethanolamine resulting in the production of acetyl-coA. This riboswitch element thus senses the metabolic level inside the cell.

In a similar fashion the cysteine biosynthesis operon of *Clostridium acetobutylicum* is controlled in a concerted and complex manner by a 5' SAM riboswitch and, on the opposite strand, by a 3' T-box that inhibits the expression of a regulatory antisense small RNA [43].

AN ANCIENT YET MOBILE REGULATORY SYSTEM

Transcription attenuation is an extremely common regulatory strategy in many bacterial species (Figure 2). Attenuation systems are often conserved among orthologous genes of evolutionarily distant species, even after associated operons have been shuffled by evolution. They are associated to a number of housekeeping operons (Table 1), and they are specifically involved in nucleic acid metabolism, which is an intriguing issue in a RNA world perspective. Furthermore, some attenuators may predate the archae/bacteria divergence as they are found conserved across both domains [2], although a later horizontal gene transfer remains conceivable. This broad phylogenetic spectrum and the functions of regulated genes argue for an ancient origin of the attenuation mechanism that was then maintained,

developed, duplicated, replaced or completed differently during the history of bacterial families.

However ancient and widespread, attenuation systems can be replaced in orthologous genes by other systems such as transcription factors responding to the same signal. Examples of such replacements include the attenuation mechanisms used for the genes of pyrimidine metabolism previously cited [40] and a number of regulatory networks, such as those for methionine [44] or tryptophan [45] metabolism. The different classes of attenuation systems are not evenly represented in all bacterial families. Generally, Firmicutes tend to use transcription attenuation whereas other bacterial phyla preferentially use translational attenuation [35]. More specifically, Firmicutes tend to use riboswitches more often than do other bacterial groups [35] (Figure 2), and the distribution of T-boxes [23] indicates that they probably arose early in a common ancestor of the Firmicutes/Chloroflexi/Deinococcus-Thermus/Actinobacteria group, and later spread by horizontal transfers, both inside the Firmicutes and in other phyla.

T-boxes represent an interesting substrate for the diversification of ligand specificity, since this specificity almost exclusively lies on the “anti-anti-codon” specifier. This suggests that T-box dissemination was achieved by duplication, transfer to other operons and specificity change by introducing a few point mutations. With regards to the peptide leader mechanism, its widespread distribution (peptide leaders are found not only in Proteobacteria but also in Bacteroidetes/Chlorobi and Thermotogales) suggests an ancient origin [24] (Figure 2). On the other hand, some replacement events involving attenuation systems seem to be far more recent. In Lactobacilli for instance, some SAM riboswitches are replaced by T-boxes [11], an event that is exclusively observed in this phylum.

Horizontal genetic transfer may represent the key mechanism for the dissemination of attenuation systems, as suggested for riboswitches, for which there are many isolated occurrences observed in certain clades [35]. The characteristic ‘stand alone’ nature of such systems, shared with other RNA-based systems (characterized by their lack of requirement of any additional protein for their function) facilitate their spreading by horizontal transfer. But recent horizontal transfers did not appear to have an equal impact on all riboswitch families. While an entire transcriptional unit containing an AdoCbl

riboswitch was transferred from a Firmicutes species to Enterobacteria, there is no evidence of such event for the lysine aptamer [35]. Another interesting trend is the seemingly independent mobility of aptamers and expression platforms in riboswitches [35] and T-boxes [23]. The majority of T-boxes, for instance, are found in Firmicutes where they control premature transcription termination, while the few cases found in Actinobacteria presenting a similar aptamer structure operate with a Shine-Dalgarno sequestration.

The role of transposable elements in the horizontal dissemination of attenuator regions is an interesting hypothesis. First, many transposable elements present an AT-rich insertion site (see e.g. [46]), and as a consequence may significantly target terminator T-tails [47] and duplicate them by their own transposition. Second, many insertion sequences encompass a Rho-independent terminator in their 5'-extremity, that prevents them from being expressed ‘constitutively’ with the gene they are inserted in, thus opening the possibility of transposons exaptation for the creation of new attenuators [48, 49].

ATTENUATORS AS EVOLUTIONARY GATEWAYS

Attenuation systems are attractive subjects for studies on early evolution since they involve RNA and can interact with small molecules in a variety of ways. However, another property of attenuators that make them even more fascinating is that they provide clues for possible evolutionary transitions from one regulatory mechanism to another.

One such transition could be from a terminator stem-loop to a transcription factor binding element. A terminator stem corresponds to a palindromic sequence on DNA and, as such, can be targeted by protein factors. For example, protein ZraP of the ZraS-ZraP TCS of *E. coli* binds a DNA palindromic region in response to high concentrations of Zn^{2+} and Pb^{2+} [50]. Intriguingly, a T-tail in this motif first led us to detect it as an intrinsic terminator. We may hypothesize that such a DNA element could be the remnant of an ancestral terminator-based system. A number of other transcription factors also bind to inverted repeats in the DNA, including members of the LysR family [51] of which genes, interestingly, often contain attenuator structures [2]. In addition, certain RNA-binding factors interact directly via

RNA hairpins [29, 52] (Figure 1e). An interesting case of pleiotropy of terminators has been shown in *Neisseria* and *Pasteurellaceae*, where terminator hairpins frequently correspond to exogenous DNA uptake signal sequences (USSs) [32]. USSs are highly conserved short (9–11 nt) sequences that occur repeatedly in the genome and facilitate the incorporation of exogenous DNA into the genome of naturally transformable bacteria. Co-opted TSSs in these species may now serve as terminators.

A second possible transition is from an amino acid operon leader to a T-box, or the converse. A leader sequence is made of codon triplets, and a specific triplet binds the tRNA anticodon in the T-box. In other words, both classes of attenuators can be modeled as having a simple codon + terminator architecture, which suggests possible evolutionary transitions between the two forms. Such transitions are not straightforward though, as actual leader sequences include a full ORF with several identical codons.

A third, even more speculative, transition is from a riboswitch amino-acid-binding motif to a T-box or an amino-acid operon leader. This hypothesis derives from the works of Yarus [53] supporting the existence of a specific affinity between amino acids and sequences enriched in cognate codon. This suggests an evolutionary pathway from a riboswitch-like aptamer that would bind amino acids through one or more codon sequences, to a T-box-like aptamer where one codon sequence is co-opted in the specifier loop. A run of codons in the riboswitch aptamer could also allow transition towards a peptide leader. Obviously, the ancient origin of these events, if they ever occurred, severely limits such reasoning.

CONCLUSIONS

Transcriptional attenuation appears to be an ancient regulatory mechanism that combines the use of a premature Rho-independent termination site with the involvement of a sensing element. The plasticity and chemical properties of RNA confer on attenuators a rich potential of responses to stimuli such as temperature change or the presence of small molecules, RNAs or proteins. Attenuators are specialized in regulating major biosynthesis pathways and certain housekeeping functions and most often act by repressing gene expression upon direct sensing of

their own product or one of the pathway metabolites. The absence of attenuation systems in eukaryotic species could find an explanation in their cellular ultrastructure, in which the genome is separated from most of the cell metabolites by the nuclear envelope. This requires that cytoplasmic conditions are processed beforehand and transmitted indirectly to the nucleus, which occurs mostly through protein-based pathways.

In addition to this functional versatility, terminator-based regulatory elements show striking evolutionary flexibility and lability. Even if certain attenuator classes are established within specific phyla and gene families (for instance T-boxes flanking aminoacyl tRNA synthetases in Firmicutes), a number of recent horizontal transfers of attenuators (e.g. a riboswitch moving from one operon to another) and exchanges between attenuators from different classes (e.g. between a riboswitch and a T-box) are observed. This lability renders the Rho-independent termination motif of particular evolutionary interest. It seems that bacterial organisms are efficiently exploiting the functional and evolutive capacities of attenuators, moving and combining them around a wide array of target genes and operons. This diversity of regulatory combinations involving attenuators appears much larger than biologists' abilities to properly annotate and characterize these systems. This untapped area of the regulome holds great potential for biotechnology applications, due to the primary importance of the gene functions involved.

SUPPLEMENTARY DATA

Supplementary data are available online at http://ma.igmors.u-psud.fr/suppl_data.

Key Points

- Attenuators combine an ancient, simple regulatory structure to a large variety of sensing mechanisms.
- Attenuators regulate key biosynthesis pathways and housekeeping functions.
- Evolutionary interchanges between different attenuation systems are observed in certain gene families.
- The majority of attenuators present in sequenced genomes are not yet characterized or even annotated.

References

1. Yanofsky C, Platt T, Crawford IP, *et al.* The complete nucleotide sequence of the tryptophan operon of *Escherichia coli*. *Nucleic Acids Res* 1981;**9**:6647–68.
2. Merino E, Yanofsky C. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet* 2005;**21**:260–4.
3. Nudler E, Gottesman ME. Transcription termination and anti-termination in *E.coli*. *Genes Cells* 2002;**7**:755–68.
4. Neal RJ, Chater KF. Bidirectional promoter and terminator regions bracket *mmr*, a resistance gene embedded in the *Streptomyces coelicolor* A3(2) gene cluster encoding methylenomycin production. *Gene* 1991;**100**:75–83.
5. Unniraman S, Prakash R, Nagaraja V. Conserved economics of transcription termination in eubacteria. *Nucleic Acids Res* 2002;**30**:675–84.
6. Narberhaus F, Waldminghaus T, Chowdhury S. RNA thermometers. *FEMS Microbiol Rev* 2006;**30**:3–16.
7. Waldminghaus T, Gaubig LC, Narberhaus F. Genome-wide bioinformatics prediction and experimental evaluation of potential RNA thermometers. *Mol Genet Genomics* 2007;**278**:555–64.
8. Chowdhury S, Maris C, Allain FH, *et al.* Molecular basis for temperature sensing by an RNA thermometer. *EMBO J* 2006;**25**:2487–97.
9. Uppal S, Rao Akkipeddi VSN, Jawali N. Posttranscriptional regulation of *cspE* in *Escherichia coli*: involvement of the short 5'-untranslated region. *FEMS Microbiol Lett* 2008;**279**:83–91.
10. Bae W, Xia B, Inouye M, *et al.* *Escherichia coli* CspA-family RNA chaperones are transcription antiterminators. *Proc Natl Acad Sci USA* 2000;**97**:7784–9.
11. Vitreschak A, Rodionov D, Mironov A, *et al.* Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet* 2004;**20**:44–50.
12. Coppins L, Hall KB, Groisman EA. The intricate world of riboswitches. *Curr Opin Microbiol* 2007;**10**:176–81.
13. Henkin TM. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* 2009;**22**:3383–90.
14. Dambach MD, Winkler WC. Expanding roles for metabolite-sensing regulatory RNAs. *Curr Opin Microbiol* 2009;**12**:161–9.
15. Sudarsan N, Lee ER, Weinberg Z, *et al.* Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 2008;**321**:411–13.
16. Regulski EE, Moy RH, Weinberg Z, *et al.* A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol Microbiol* 2008;**68**:918–32.
17. Dann CD, Wakeman CA, Sieling CL, *et al.* Structure and mechanism of a metal-sensing regulatory RNA. *Cell* 2007;**130**:878–92.
18. Serganov A, Huang L, Patel DJ. Structural insights into amino acid and gene control by a lysine riboswitch. *Nature* 2008;**455**:1263–7.
19. Garst AD, Heroux A, Rambo RP, *et al.* Crystal structure of the lysine riboswitch regulatory mRNA element. *J Biol Chem* 2008;**283**:22347–51.
20. Sudarsan N, Hammond MC, Block KF, *et al.* Tandem riboswitch architectures exhibit complex gene control functions. *Science* 2006;**314**:300–4.
21. Kim JN, Breaker RR. Purine sensing by riboswitches. *Biol Cell* 2008;**100**:1–11.
22. Gutiérrez-Preciado A, Henkin TM, Grundy FJ, *et al.* Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. *Microbiol Mol Biol Rev* 2009;**73**:36–61.
23. Vitreschak AG, Mironov AA, Lyubetsky VA, *et al.* Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA* 2008;**14**:717–35.
24. Vitreschak AG, Lyubetskaya EV, Shirshinet MA, *et al.* Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiol Lett* 2004;**234**:357–70.
25. Lyubetskaya EV, Leont'ev LA, Gelfand MS, *et al.* Search for alternative RNA secondary structures regulating expression of bacterial genes. *Mol Biol* 2003;**37**:707–16.
26. Nomura M, Yates JL, Dean D, *et al.* Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: structural homology of ribosomal RNA and ribosomal protein mRNA. *Proc Natl Acad Sci* 1980;**77**:7084–8.
27. Draper DE. How do proteins recognize specific RNA sites? New clues from autogenously regulated ribosomal proteins. *Trends Biochem Sci* 1989;**14**:335–8.
28. Romby P, Springer M. Bacterial translational control at atomic resolution. *Trends Genet* 2003;**19**:155–61.
29. Bonner ER, D'Elia JN, Billips BK, *et al.* Molecular recognition of *pyr* mRNA by the *Bacillus subtilis* attenuation regulatory protein PyrR. *Nucleic Acids Res* 2001;**29**:4851–65.
30. Amster-Choder O. The *bgl* sensory system: a transmembrane signaling pathway controlling transcriptional antitermination. *Curr Opin Microbiol* 2005;**8**:127–34.
31. de Hoon MJL, Makita Y, Nakai K, *et al.* Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS Comp Biol* 2005;**1**:e25.
32. Kingsford CL, Ayanbule K, Salzberg SL. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 2007;**8**:R22.
33. Konan KV, Yanofsky C. Rho-dependent transcription termination in the *tna* operon of *Escherichia coli*: Roles of the *boxA* sequence and the *rut* site. *J Bacteriol* 2000;**182**:3981–8.
34. Tsubouchi T, Mineki R, Taka H, *et al.* Leader peptide-mediated transcriptional attenuation of lysine biosynthetic gene cluster in *Thermus thermophilus*. *J Biol Chem* 2005;**280**:18511–16.
35. Barrick JE, Breaker RR. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* 2007;**8**:R239.
36. Yao Z, Barrick J, Weinberg Z, *et al.* A computational pipeline for high-throughput discovery of cis-regulatory non-coding RNA in prokaryotes. *PLoS Comput Biol* 2007;**3**:e126.
37. Merino E, Yanofsky C. Regulation by termination-antitermination: a genomic approach. In: Sonenshein AL, Hoch JA, Losick R (eds). *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*. Washington: ASM Press, 2002:323–36.

38. Griffiths-Jones S, Bateman A, Marshall M, *et al.* Rfam: an RNA family database. *Nucleic Acids Res* 2003;**31**:439–41.
39. Dufayard JF, Duret L, Penel S, *et al.* Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. *Bioinformatics* 2005;**21**:2596–603.
40. Turnbough CL, Jr, Switzer RL. Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiol Mol Biol Rev* 2008;**72**:266–300.
41. Jensen-MacAllister IE, Meng Q, Switzer RL. Regulation of pyrG expression in *Bacillus subtilis*: CTP-regulated antitermination and reiterative transcription with pyrG templates in vitro. *Mol Microbiol* 2007;**63**:1440–52.
42. Fox KA, Ramesh A, Stearns JE, *et al.* Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in *Enterococcus faecalis*. *Proc Natl Acad Sci* 2009;**106**:4435–40.
43. André G, Even S, Putzer H, *et al.* S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res* 2008;**36**:5955–69.
44. Rodionov DA, Vitreschak AG, Mironov AA, *et al.* Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems. *Nucleic Acids Res* 2004;**32**:3340–53.
45. Merino E, Jensen RA, Yanofsky C. Evolution of bacterial trp operons and their regulation. *Curr Opin Microbiol* 2008;**11**:78–86.
46. Devalckenaere A, Odaert M, Trieu-Cuot P, *et al.* Characterization of IS1541-like elements in *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. *FEMS Microbiol Lett* 1999;**176**:229–33.
47. Odaert M, Devalckenaere A, Trieu-Cuot P, *et al.* Molecular characterization of IS1541 insertions in the genome of *Yersinia pestis*. *J Bacteriol* 1998;**180**:178–81.
48. Davis MA, Simons RW, Kleckner N. Tn10 protects itself at two levels from fortuitous activation by external promoters. *Cell* 1985;**43**:379–87.
49. Beuzon CR, Marques S, Casades J. Repression of IS200 transposase synthesis by RNA secondary structures. *Nucleic Acids Res* 1999;**27**:3690–3695.
50. Leonhartsberger S, Huber A, Lottspeich F, *et al.* The hydH/G genes from *Escherichia coli* code for a zinc and lead responsive two-component regulatory system. *J Mol Biol* 2001;**307**:93–105.
51. Schell MA. Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol* 1993;**47**:597–626.
52. Babitzke P, Romeo T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol* 2007;**10**:156–63.
53. Yarus M, Caporaso JG, Knight R. Origins of the genetic code: the escaped triplet theory. *Annu Rev Biochem* 2005;**74**:179–98.