Alternative transcriptional regulation in genome-reduced bacteria
Samuel Miravet-Verde, Verónica Lloréns-Rico and Luis Serrano

Introduction
Genome-reduced bacteria are of remarkable interest as model organisms to study basic aspects of bacterial physiology. Because of their inherent simplicity, they are attractive for systems biology studies, whose results can be generalized to larger, more complex bacteria. These organisms have encountered defined niches to colonize as endosymbionts or pathogens, and have adapted to their environments by eliminating genes that are not required for their development. For instance, they have usually lost metabolic pathways to synthesize elements present in their natural environment [1]. Also, this niche adaptation has affected how gene expression is regulated in these organisms. Transcription factors (TFs), which have been traditionally considered the major drivers of transcriptional regulation, are scarce in bacteria with small genomes. In bacterial models like Escherichia coli or Bacillus subtilis, TFs represent 5–6% their total number of genes. This number is reduced by half (2.5% on average) in the Mollicutes class, a bacterial group including multiple minimal bacteria, most of them Mycoplasmas [2*]. A comparative analysis of 50 Mollicutes genomes identified 1–5 global regulators and up to 15 TFs in the Mycoplasmas with larger genome sizes [2*]. However, to the best of our knowledge, none of the putative global regulators has been characterized with the exception of the housekeeping sigma factor. Known transcription factors, including an additional sigma factor [3], only regulate a handful of genes [4].

Despite the tiny repertoire of TFs, these bacteria have not lost the ability to respond to a variety of external perturbations [4]. Therefore, it is possible that novel TFs remain undiscovered given the percentage of genes with unknown functions in these organisms, or that non-TF proteins with moonlighting functions act as TFs. Alternatively, different forms of regulating gene expression must exist, and may prevail, in these organisms. These alternative regulatory elements are probably not unique to genome-reduced bacteria, but they become more important as the process of genome reduction removes TFs to minimize the DNA content in these organisms. These alternative mechanisms of gene regulation are probably ancestral, as they are based in the chromosome structure and/or the intrinsic DNA or RNA sequences and not in proteins. The regulation they confer could have a smaller dynamical range and is more subtle than that by transcription factors, which makes it hard to observe in more complex bacteria. In this review, we focus on these other regulatory elements, from genome-wide to transcript-specific.

Genome structure and DNA topology
First high-resolution 3D structure of a bacterial chromosome, obtained for Caulobacter crescentus, showed 23 interacting regions ranging from 30 to 400 kb bounded by
highly transcribed genes, known as chromosomal interaction domains (CIDs) [5]. Lately, ~20 CIDs were defined in *Bacillus subtilis* with a size between 50 and 300 kb [6]. Disposition of these elements is regulated by DNA supercoiling, which is controlled by topoisomerases [7] and nucleoid-associated proteins (NAPs) [8] (Figure 1a). *B. subtilis* presents four DNA topoisomerases: two ATP-independent (I and III) and two ATP-dependent (II, known as DNA gyrase, and IV) [9]. Minimal cells commonly present no topoisomerase III and a significant reduction of NAPs [10,11]. With such a low number of DNA-binding proteins it was questionable whether small bacteria would preserve a chromosomal organization. A recent study in *Mycoplasma pneumoniae* found that small bacteria have enough components to maintain a defined chromosome structure and the presence of CIDs. In addition, this study provides the first evidence that genes inside CIDs tend to be co-regulated but the underlying mechanism to achieve this remains unknown. Interestingly, CIDs in *M. pneumoniae* are smaller (15–33 kb) but more frequent (44 CIDs) than *C. crescentus* and *B. subtilis* [11]. Additionally, promoters are sensitive to local superhelical state as it regulates the distance between the elements participating in the promoter [12]; even in small-genome bacteria with reduced number of topoisomerases [13]. Yus et al., in preparation). Finally, ATP controls the ratio of ATP dependent/independent topoisomerases with direct effect on supercoiling and could imply a regulatory link between metabolism and genome topology and, consequently, expression [13].

**Genome organization in operons**

Genome organization in operons constitutes a first level of gene regulation in prokaryotes. As transcription and translation occur simultaneously in bacteria, positional effects exist, and expression levels of the individual proteins in an operon are inversely proportional to the distance to the transcription initiation site of the operon [14]. This represents a level of regulation that is used not only in small but in all bacteria.

Traditionally, operons have been treated as static entities. However, recent research has shown that these structures are highly dynamic, being able to adapt in response to changing conditions, mainly thanks to termination, generating large transcripts or super-operons in some conditions, while producing short transcripts of sub-operons in others (Figure 1b) [15]. In *M. pneumoniae*, this condition-dependent transcriptional read-through can explain a large part of how transcription is regulated [15]. This mechanism has been shown to occur also in larger bacteria such as *E. coli* and *B. subtilis* [16].

**Bacterial promoters and transcription initiation**

Promoter regions require certain features that make them recognizable by the RNA polymerase (RNAP) and the different TFs. Besides specific motifs binding sites for TFs, the most important sequence features are the boxes recognized by the RNAP complex and the different sigma factors. The housekeeping sigma factor binds two regions: the −10 box or Pribnow motif, and the −35 box. In genome-reduced bacteria, promoters have evolved towards the elimination of the −35 box, as this is non-existent or highly degenerated (Figure 1b) [8,17,18]. In *Buchnera aphidicola*, an aphid symbiont with a minimal genome, regions similar to the −10 box of *E. coli* have been found, while a −35 motif has been only found upstream the rRNA genes [19]. In Gram-positive bacteria like *B. subtilis*, absence of a −35 element has been shown to be compensated if the Pribnow motif is preceded by a ‘TG’ dinucleotide (the so-called extended −10 box), but this short motif is present in only a handful of promoters in *Mycoplasma gallisepticum* [17] and is not essential in determining promoters in *M. pneumoniae* [18]. This reduction in promoter complexity could be due to the scarcity of alternative sigma factors. This raises a question as to what makes promoters determine initiation of transcription and recognition by the RNAP complex. A recent study in *M. pneumoniae* points to the importance of the bases immediately surrounding the Pribnow motif, which tend to be A/T rich [20].

The structure of these regions is also important to trigger transcription. The double-stranded DNA should be less stable at the promoter region to unwind and accommodate the RNAP complex. Although the unwinding of the double helix is energetically favored at the promoters, the open complex formed between the promoter and the RNAP can be unstable. Unstable complexes require high concentrations of the initiating NTP (iNTP) to be stabilized so that RNA synthesis can be launched immediately. Otherwise, these complexes rapidly dissociate and transcription initiation is not produced. In contrast, very stable complexes require lower concentrations of the iNTP, as they will not easily dissociate [21]. Later, it was shown that the +2 nucleotide also modulates transcription initiation [22]. This mechanism establishes a link between cellular metabolism and transcriptional regulation and is not unique of genome reduced bacteria, but in the absence of major regulators this might be an elegant way to coordinate the expression of large groups of transcripts with identical +1 and +2 bases. An example of this nucleotide-based regulation includes the response to amino acid starvation (stringent response) in *B. subtilis*. In this scenario, concentration of ATP increases while GTP decreases as a consequence of the synthesis of (p)ppGpp (Figure 1b) [23]. Upregulated genes in this condition have adenosine in the +1 position, while downregulated promoters have guanosine. This effect could also be present and play a major role in the absence of many TFs in minimal bacteria as a regulatory mechanism dependent only in sequence composition.
Figure 1

(a) Genome-level regulation

NAPs
Topoisomerases
CID
[ATP]

Topoisomerases

Topo II (DNA gyrase)
Topo I, topo IV

Topo II (DNA gyrase), topo IV

Positive supercoiling

(b) Operon-level regulation

Transcription initiation

TANAAT
Promoter
iNTP

Transcription termination

Condition A

Condition B

Terminator readthrough

(c) RNA-level regulation

RNases J1/J2
RNase E
REPs
RNase III
RNase Y

RNase R

Riboswitches

Condition X
Condition Y

Termination hairpin formed

sRNAs (trans-encoded)

RNases recruitment

Ribosome recruitment

(+)

(−)

(+)

(−)

TF-independent regulation of transcription at three different levels observed in genome-reduced bacteria. (a) Genome-wide level. At this level the principal actuators are the genome structure, organized in chromosomal interaction domains (CIDs) and maintained by nucleoid-associated proteins (NAPs); and the supercoiling, regulated by gyrases and topoisomerases (topo). ATP affects supercoiling through the regulation of the
Termination
Transcription termination in bacteria can be accomplished by rho-dependent or intrinsic termination (IT). First type involves Rho protein moving through the nascent RNA and disassembling the transcription machinery [24]. IT depends on terminator sequences composed of a stem-loop hairpin followed by a poly-uridine (poly-U) tail. Poly-U induces the RNAP backtracking towards the nearest hairpin that disintegrates the elongation complex [25]. Rho is usually essential in Gram-negatives but Gram-positive model cells are viable without it [26,27]. Remarkably, few species directly lack this gene and any homolog: *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Mycoplasma genitalium*, *M. pneumoniae*, *Ureaplasma urealyticum*, and *Synechocystis* sp. PCC6803 [28]. Species in this group are all Gram-positive, present low GC contents and, except *Synechocystis* sp., have genome sizes between 0.5 and 2 Mb. Interestingly, low GC content has been presented as an impediment to form stable terminators. Analysis carried in several prokaryotes, including Rho-lack Mycoplasmas, showed that no free energy minimum to form hairpins is achieved close to stop codons although termination still occurs [29]. This could imply the existence of a third unknown mechanism that could be specially relevant in Rho-lack organisms, which are most of them genome-reduced bacteria.

IT regulation mainly relies on hairpin stability and poly-U length but three additional elements need to be considered. Firstly, low uridine triphosphate (UTP) concentration helps termination [30]. Secondly, elongation factors modify RNAP processivity and its sensitivity to terminators and they are reduced in minimal cells, like NusG or NusB, inexistent in most of them [10,31]. Finally and as mentioned above, IT can be condition-dependent as it has been observed cases of readthrough and imperfect termination as response to different environmental stimuli [15*,16,32] (Figure 1b).

Riboswitches
Riboswitches are segments within an mRNA that bind metabolites triggering a structural change that affects the encoded protein expression. This effect is a direct consequence of hiding or exposing terminators or ribosome binding sites [33]. Affecting transcription, only ribo-regulation based on termination has been defined with riboswitches usually located within 5’-UTR regions of metabolic genes and controlled by metabolite ligands appearing in the same pathways of the genes they regulate [34]. When active, they transform an anti-terminator in an intrinsic terminator, producing a premature termination, or vice versa producing readthrough (Figure 1c). Multiples cases of this ribo-regulation have been defined with strong importance in bacterial physiology and virulence [35*]. Despite knowledge about ribo-regulation in small bacteria is still narrow, they are good alternatives to regulate genes in a TF-independent manner and, as occurs in termination, saving genomic space with a mechanism embedded into the sequence itself. As example, we know that multiple metabolic pathways in minimal cells are reduced to the core as they receive multiple resources from the host they parasite and some of these pathways commonly include regulation by riboswitches [36,37]. More interestingly, there are cases where ribo-regulation in small bacteria has evolved to high levels of complexity. One example includes multiple variants of guanine riboswitches found in the genome-reduced bacterium *Mesorhizobium loti* that are not seen in other organisms [38].

Small RNAs
Non-coding or small RNAs (sRNAs) in bacteria have traditionally been thought to act as gene expression regulators, either at the transcriptional, post-transcriptional or translational level [39–41]. The 6S RNA, which directly regulates the activity of the RNAP, is found only in Rickettsias, but not in other genera of genome-reduced bacteria such as *Buchnera* or *Mycoplasma* [42]. Despite the variety of possible mechanisms of action described for sRNAs (Figure 1c), only a minority of the discovered sRNAs have been characterized, most of which correspond to the trans-encoded sRNAs located in intergenic regions. In some *Mycoplasma* species, different intergenic sRNAs and their targets have been annotated using in silico approaches, and some of them have been found to be transcriptionally regulated differentially in various conditions [43]. In *Rickettsia conorii*, interaction between an intergenic sRNA and its targets could be experimentally validated [44]. However, genome compaction in small bacteria has caused intergenic regions to shrink substantially, therefore reducing the number of trans-sRNAs [45**]. In contrast, genome-reduced bacteria are rich in cis-encoded antisense sRNAs. The functions of these has not been studied in depth, but a recent study provides evidence that in *M. pneumoniae*, most antisense RNAs could be the product of pervasive transcription arising at spurious
promoters [45**]. The low information content of promoters in these organisms, probably associated to the decrease of sigma factors, together with the higher probability of mutations from G/C to A/T in bacteria [46] could allow for a rapid formation of novel functional promoters, giving rise to these non-coding transcripts. Indeed, the number of antisense transcripts in bacteria correlates with the AT content of their genomes [45**]. Although this suggests that the individual antisense RNAs do not have a regulatory function, this pervasive transcription could have a role in generating variability in the bacterial population, and probably this phenomenon is not unique to genome-reduced bacteria. Other class of small, non-coding RNAs are TSS-associated RNAs, that have been found in Mycoplasmas [47] and have been hypothesized to prevent transcription elongation until the correct RNAP complex has been assembled.

**Post-transcriptional regulation**

Maturation and degradation are essential events controlling RNA concentration catalyzed by enzymes with different specificities between stable (rRNA, tRNA) and messenger RNA (mRNA) including: exoribonucleases digesting from one end of the molecule and endoribonucleases cleaving the RNA internally [48]. Degradation and maturation start with an endoribonucleolytic primary cleavage as exoribonucleases cannot target newly produced RNA [49]. At this level, Gram-positive minimal cells do not show a clear reduction and they conserve the most important endoribonucleases found in *B. subtilis*: RNases III, H, J1, J2, P, Y and NrnA [50**]. After a cleavage, stable and mRNA can be digested from 3′-end by PNPase, RNase R, YhaM, YhcR and Yvaj exoribonucleases in *B. subtilis* [51]. From 3′-end, RppH can remove the phosphate protection and trigger a rapid degradation by RNase J1 acting as 5′-exoribonuclease [51]. Unlike endoribonucleases, minimized bacteria show a strong reduction in 3′ exoribonucleases with only RNase R conserved. In addition, lack of RppH and not reported role of J1 as exoribonuclease in small bacteria make unlikely their 5′-exoribonuclease activity [50**, 51–53] (Figure 1c).

In RNA maturation RNases III and YhaM (participants in degradation too) and specific RNases Bsn, M5 and PH complete the task in *B. subtilis* [54]. No specific enzymes have been found in small Gram-positive bacteria and degradation-related RNases (III, P and R) participate in maturation of stable RNA [55].

In addition, ribonucleases can interact in a complex specialized in degradation and processing called degradosome. Gram-positive degradosome consists of three RNases (J1, J2 and Y), PNPase, CshA (RNA helicase) and two glycolytic enzymes (enolase and phosphofructokinase) [56]. Degradosome in small cells remains undefined and lack of PNPase is an important limitation; however, either as a remanent or because degradosome exists in small bacteria, same glycolytic enzymes than in *B. subtilis* interact in *M. pneumoniae* [57].

**REP elements**

Repetitive Extragenic Palindromic (REP) elements are species-specific conserved sequences that form RNA secondary structures. These elements were first found in *E. coli* representing close to 1% of its genome [58]. Lately, these have been characterized in minimal cells with multiple examples in *Myoplasma* spp. [59*].

Regulatory spectrum of REP sequence has not be fully explored yet but they seem to act at many levels during transcription. Firstly, REP elements preferentially bind gyrase so their effect could be extended to affect DNA supercoiling with its respective impact in transcription regulation [60]. Secondly, their recurrence within intercistronic regions has been associated to regulation of relative expression of genes within the same operon [61]. Finally, they can interact with the degradation machinery as potential target of RNaseIII and, due to their common presence upstream to terminators in the 3′-untranslated region (3′-UTR), it has been suggested REP elements could protect mRNA against 3′–5′ exoribonucleolytic activity [48,58].

**Conclusion**

Minimal bacteria arisen by degenerative evolution have in common a significant reduction of the number of proteins encoded within their genomes. This reduction implies a lack of multiple TFs resulting in an increased relevance of TF-independent transcriptional regulation at genome-wide, operon and transcript levels.

At the genome-wide level, minimal cells have conserved a minimal set of proteins to maintain a structured genome and to control its superhelical state, both with direct effect on transcription. In addition, supercoiling could have an extended regulatory role in genome-reduced bacteria, as the high level of compaction makes it more likely to affect several operons with single local adjustment of the superhelical density.

A second level based on operon organization comprises regulation of transcription initiation and termination. At this level we observe the impact of DNA sequence composition in transcriptional regulation with promoter motifs and the effect of the iNTP on initiation. We also find phenomena such as transcriptional read-through, with special relevance in minimal bacteria due to their high degree of compaction. Last type of TF-independent regulation occurs at the transcriptional level, where RNA structures that are part of the mRNA itself or additional interacting RNAs could critically impact the functional RNA concentration. For instance, riboswitches are a good alternative to regulate metabolic genes and operons based
on the structure of the 5’ end of the mRNA. Transcription-level regulation also includes degradation of the mRNA where REP sequences (mRNA itself) and additional sRNAs interacting with it participate in the recruitment of ribonucleases, controlling the mRNA availability.

Throughout this review, we have noted that besides TFs, the specific protein machinery of these alternative regulatory elements has also been reduced (NAPs, elongation and termination factors and exoribonucleases have been lost in these organisms). Nevertheless, the core functionality remains, as cells are capable of displaying different transcriptional responses to perturbations. This functionality could thus rely on single proteins with moonlighting functions (e.g. RNase R) or more interestingly, on mechanisms that are implicit to the genome features or RNA molecules themselves without requiring any encoded protein. We believe that these could be ancestral mechanisms, that are not unique to genome-reduced bacteria, but that can be observed and studied in these organisms because the lack of TFs makes them more relevant. A question that remains open is to which extent each of the alternative mechanisms is responsible for the RNA regulation inside the cell, that is, how much of the RNA dynamics can be explained by each of these elements. Currently, there is no framework that allows to integrate the effect of the different mechanisms, but recent advances in modelling approaches, such as multi-scale models or even whole-cell models, could shed light on this question.

Conflict of interest
Nothing declared.

Acknowledgements
We thank Drs. Eva Yus and Dr. Maria LLuch-Senar for providing valuable comments to the manuscript. We acknowledge support of the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program under agreement No 670216 (MYCOCHASSIS) and the Spanish Ministry of Economy and Competitiveness, ‘Centro de Excelencia Severo Ochoa 2013-2017’.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


To date, this is the most comprehensive study of genome topology in a minimal cell and its impact in transcription.


First article describing that transcriptional read-through occurs at a genome-wide scale in a genome-reduced bacterium.


In this study, authors analyze in depth the promoter features in a genome-reduced bacterium altogether with a precise mapping of the 5’ ends of all the RNAs. They pinpoint the importance of the 190 box surrounded by AT-rich regions, and the preference for A/G at the +1 position of the transcript. Besides, they also study the features of transcription termination signals.


This article presents a novel promising method to study implication of ribo-regulation based on transcription across different bacterial species.


In this study, the authors provide evidence that antisense RNAs, a type of small RNAs, can be the product of transcriptional noise in genome-reduced bacteria, and they suggest that this might be true for all bacterial species.


This article presents the distribution and possible roles of REP sequences in bacteria with special focus in reduced organisms.
