Post-transcriptional regulation of gene expression in the human opportunistic pathogens of the Burkholderia cepacia complex

Christian G. Ramos¹,²*, André M. Grilo¹*, Silvia A. Sousa¹,², Joana R. Feliciano¹, Jorge H. Leitão¹,²*

¹) Institute for Biotechnology and Bioengineering; ²) Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, Torre Sul, Piso 6, 1049-001, Lisboa, Portugal
*correspondência: jorgeleitao@tecnico.ulisboa.pt; christian.ramos@tecnico.ulisboa.pt

Regulation at the post-transcription level is of critical importance for the fast adaptation of bacterial pathogens to their host environment. Small non-coding RNAs and the RNA chaperone Hfq are major players in this fast adaptation. Here we briefly present mechanisms used by sRNAs and Hfq to regulate gene expression at the post-transcription level, together with results from our research on sRNAs and Hfq-like proteins from Burkholderia cepacia complex bacteria.

Introduction
In order to succeed, bacterial pathogens have evolved many strategies to survive the host defences and to express their virulence factors. This requires a tight regulation of gene expression to rapidly adapt to the changing conditions faced by the pathogen in the host environment. Our understanding of the mechanisms involved in this adaptation has risen enormously, and until recently, research has been focused on the regulation of gene expression at the transcription level. A wide range of detailed mechanisms, including gene expression activators and repressors, and associated complex regulatory networks have been described.

In the last years, accumulating evidence pointed out regulation at the post-transcription level as playing a prominent role in the fast adaptation of bacterial pathogens to their host environment. The main post-transcription regulation mechanisms described so far include changes in the protein levels and activity, and the fast adjustment of the levels of relevant mRNAs. In the present work we will focus on the fast adjustment of mRNA levels by mechanisms involving small non-coding RNAs (hereafter designated sRNAs) and the RNA chaperone Hfq, a small abundant protein that mainly acts as a mediator of RNA-RNA interactions.

Examples will be presented on the physiological roles and regulatory pathways involving sRNAs and RNA chaperones from bacteria of the Burkholderia cepacia complex, a group of pathogens that can cause severe and often lethal infections of particular concern among cystic fibrosis patients.

Bacterial sRNAs
When it was first described in 1984, MicF was the sole example of a bacterial RNA acting at the post-transcription level, controlling the translation of a specific mRNA by an antisense mechanism (Mizuno et al., 1984). Since then, a plethora of bacterial sRNAs have been identified, mainly due to advances in high-throughput sequencing technologies such as deep-sequencing, together with the development of powerful bioinformatics tools. Bacterial sRNAs can be characterized as untranslated regulatory RNAs (they do not encode proteins) typically ranging 50-400 nucleotides in length, often encoded in intergenic regions. Bacterial sRNAs, as well as eukaryotic non-coding RNAs, act to modulate gene expression at the post-transcription level. Most of the bacterial sRNAs exert their regulatory function by base-pairing with a short segment of a specific mRNA, therefore named target mRNA. This base-pairing can be as short as 7 nucleotides in length (the seed) and often leads to an accelerated decay of the mRNA, although there are some cases in which this interaction leads to the stabilization of the
Bacterial sRNAs can also be classified as cis- or trans-encoded. cis-encoded sRNAs are encoded in the same loci as their target mRNA, in the complementary strand. Their regulatory action is thought to be highly effective, as cis-encoded sRNAs specifically target a particular messenger, usually without the requirement of additional factors. trans-encoded sRNAs exert their regulatory action on target mRNAs encoded in loci distinct from that where they are encoded, and arguably represent the most extensively characterized class of bacterial sRNAs (Wassarman, 2002). Many trans-encoded sRNAs target multiple mRNAs as a result of the limited complementarity shared with the mRNA target. In order to interact with their targets, trans-encoded sRNAs require the activity of the RNA chaperone Hfq, a small protein that mediates the interactions of sRNAs with their targets (Vogel and Luisi, 2011). Despite their classification as trans-encoded or cis-encoded sRNAs, both classes of bacterial sRNAs act in trans by base-pairing in a shorter or longer extension with their mRNA targets, often with the help of the RNA chaperone Hfq (Wagner et al., 2002). This interaction can lead to 1) the occlusion of the ribosome binding site, obstructing the binding of ribosomes to the messenger and thus blocking protein synthesis, and is often accompanied with the active recruitment of RNase E and consequent accelerated degradation of the RNA duplex (Fig. 1A); 2) the restructuring of the secondary structure of the mRNA, exposing the otherwise occluded ribosome binding site and thus allowing the binding of ribosomes to the mRNA with the consequent stabilization of the messenger and activation of protein synthesis (Fig. 1B).

Another important class of bacterial sRNAs act by binding to proteins, sequestering them and thus negatively regulating their activity. This is the case of CsrB (Babitzke and Romeo, 2007) and other sRNAs, which will not be discussed in this work.

**The Hfq and Hfq2 RNA chaperones from Bcc bacteria**

Over the last years, our research interests have been focused on the unveiling of genetic determinants of virulence of bacteria of the *Burkholderia cepacia* complex (Bcc). Bcc is a group of plant, animal and human opportunistic pathogens, distributed by at least eighteen closely related bacterial
species of the β proteobacteria, widely distributed by both natural and man-made habitats (Peeters et al., 2013). These bacteria are particularly problematic to patients suffering from the genetic diseases cystic fibrosis or chronic granulomatous disease (Govan and Deretic, 1996). More recently, Bcc emerged as important pathogens among hospitalized patients, particularly oncological patients (Leitão et al., 2010; Mahenthiralingam et al., 2008). These bacteria are particularly feared due to the easy patient-to-patient transmission of epidemic strains, their intrinsic resistance to the many clinically-available antibiotics rendering their eradication virtually impossible, and their ability to cause the “cepacia syndrome”, a rapid and fatal necrotizing pneumonia (Drevinek and Mahenthiralingam, 2010; Leitão et al., 2008; Leitão et al., 2010; Mahenthiralingam et al., 2005).

To identify genetic determinants from Bcc bacteria, a strategy was pursued involving the construction of a library of mutants obtained with plasposons from the clinical isolates B. cepacia IST408 and B. cenocepacia J2315, followed by selection of mutants attenuated in their virulence towards the nematode Caenorhabditis elegans used as an infection model (Sousa et al., 2011). Among the mutants recovered, one was found to harbor a plasposon insertion in a gene encoding a protein 81 % identical to the E. coli Hfq (HfqEC). The E. coli Hfq protein is composed of 102 amino acid residues, with a conserved bipartite motif composed by two regions named Sm1 and Sm2, required for RNA binding (Fender et al., 2011). The conserved core spans amino acid residues 7 to 66, forming a α-helix at the N-terminus, followed by five-stranded antiparallel β-sheets. This structure enables the protein to assemble into a quaternary ring-shaped homo-hexameric “Doughnut-like” structure, in which the inner proximal face of the Sm domain has strong affinity to A/U-rich RNAs, and the distal face exhibits a higher affinity for A-rich RNAs (Fender et al., 2011). The B. cepacia gene, hfqBC, was also found to be encoded within a genetic region with similarities to the genetic region of E. coli where the hfq gene is encoded (Sousa et al., 2010). Results from bioinformatics analysis performed to localize the gene in the B. cenocepacia J2315 reference genome (Holden et al., 2009) indicated the existence of a second gene encoding an additional hfq-like gene, named hfq2. In fact, further comparative genomics analysis led us to confirm that Bcc bacteria encode two distinct and functional hfq-like genes in their genomes, both located in chromosome 1 (Fig. 2).

![Localization of the genes hfqBC (A) and hfq2BC (B) encoding the RNA chaperones HfqBC and Hfq2BC, respectively. Numbers indicate the nucleotide position in chromosome 1 of Burkholderia cenocepacia J2315.](image)

The hfqBC gene encodes a protein composed of 79 amino acid residues, while the hfq2BC gene encodes a 188 amino acid residue. The Hfq2BC protein differs from HfqBC by presenting a 109 amino acid extension at its C-terminus, organized in 4 glycine-rich repeats. Besides Bcc bacteria, only a few other prokaryotes are known to encode 2 hfq-like genes within their genome sequences, namely the archaeal species
Methanobacterium thermaeutrotrophicum and Archaeoglobus fulgidus, and the bacterial species Bacillus anthracis, Magnetospirillum magnetotacticum, and Novosphingobium aromaticivorans (Ramos et al., 2011).

Most of the current knowledge on the structure and functions of Hfq derives from the E. coli protein (HfqEC), the first to be described (Kajitani et al., 1994). Mutations in the hfq gene of E. coli result in a pleiotropic phenotype. In agreement, mutations in the hfq gene in 4 clinical isolates belonging to Bcc also resulted in multiple phenotypes, despite the existence of two functional genes in the genomes of the bacterium (Sousa et al., 2010; Ramos et al., 2011). Mutations in the hfq2BC gene also resulted in similar phenotypes. These phenotypes included resistance to stress conditions mimicking those faced by the bacterium when infecting the respiratory tract of the human host, such as nutrient starvation, oxidative and thermal stress, exposure to acidic pH, high osmolarity, ability to form biofilms in vitro, and ability to kill the nematode C. elegans (Fig. 3).

Identification and functional analysis of sRNAs from Bcc bacteria
A strategy to experimentally isolate and identify Bcc sRNAs that require the protein HfqBC for their action was set-up, based on the in vitro co-precipitation of RNAs with purified HfqBC (Ramos et al., 2013a). For this purpose, total RNA was extracted from B. cenocepacia J2315 cells (a clinical isolate responsible for several deaths among cystic fibrosis patients, worldwide), size-fractioned, and the small-sized RNA fractions were incubated with HfqBC. The HfqBC-RNA complexes were isolated, RNA was recovered, reverse-transcribed into cDNA, cloned, sequenced and analyzed by bioinformatics tools. This strategy allowed the identification of 24 distinct sRNAs from Bcc, and their biological functions are currently being investigated (Ramos et al., 2013a).

Two additional sRNAs from B. cenocepacia
J2315 were identified and functionally characterized by our research group, MtvR and h2cR. MtvR is a 136-nucleotide long trans-encoded sRNA, located in an intergenic region, between the genes encoding a leucyl-tRNA and a putative RNase R, with homologues restricted to members of the Bcc (Ramos et al., 2013b). Northern blot analysis allowed the detection of transcripts corresponding to the sRNA in cells at different phases of the batch growth curve, being more abundant in stationary phase cells. Half-life time determinations showed that the sRNA is about 3 times more stable in the stationary phase than in the early exponential phase. More than 300 putative MtvR mRNA targets were predicted by bioinformatics tools. A total of 17 out of 19 selected mRNAs were confirmed experimentally as MtvR targets by means of silencing or overexpression of the sRNA (Ramos et al., 2013b). Among the putative targets of MtvR, hfqBC was one of the mRNAs shown to be regulated by this sRNA. Analysis of the interactions of MtvR and hfqBC mRNA revealed that the sRNA binds exclusively to the 5'-untranslated region of the messenger, resulting in the accelerated decay of hfqBC mRNA and in a lower abundance of the encoded protein, HfqBC (Ramos et al., 2013b). Remarkably, strains overexpressing MtvR or with the sRNA silenced exhibited pleiotropic phenotypes, indicating that MtvR is a global regulator of gene expression in Bcc bacteria.

The h2cR was identified as a cis-encoded sRNA from B. cenocepacia J2315, located in the reverse strand of the 5'-untranslated region of the gene encoding the RNA chaperone Hfq2BC (Ramos et al., 2012). Northern blot analysis revealed that h2cR is transcribed mainly in cells at the early exponential phase, decreasing to barely detectable levels in cells in the stationary phase. The sRNA was found to negatively affect the levels of hfq2BC mRNA, leading to decreased levels of the protein. Molecular analysis revealed that h2cR specifically binds to nucleotides -220 to -40 of the 5'-untranslated region of hfqBC mRNA. Interestingly, both the h2cR transcript and the hfq2BC mRNA were found to be stabilized by HfqBC, the other RNA chaperone from Bcc bacteria (Ramos et al., 2012). The combined regulatory action of the sRNAs MtvR and h2cR results in a feed-forward loop that tightly regulates, at the post-transcription level, the mRNAs corresponding to the RNA chaperones HfqBC and Hfq2BC, which in turn are themselves major regulators of gene expression in Bcc bacteria (Fig. 4).

![Fig. 4. Regulatory circuitry showing the action exerted by the sRNAs MtvR and h2cR on the hfqBC and hfq2BC mRNAs, respectively.](image)

**Perspectives**
Over the last few years a huge progress on the knowledge of virulence traits used by several bacterial pathogens has been achieved. The discovery of sRNAs is a hallmark in our understanding on how pathogens finely tune their gene expression to thrive in the changing environment they face when infecting their hosts. The unveiling of sRNAs and associated post-transcriptional regulatory mechanisms will most certainly empower us with new tools to rationally design novel therapeutic strategies to tackle the infections caused by bacterial pathogens.

**Acknowledgements**
This work was supported by Fundação para a Ciência e a Tecnologia (FCT, Portugal), www.spmicrobiologia.pt

References


