1 **Overlapping transcription and bacterial RNA removal**

2

3

4 Iñigo Lasa[1], Maite Villanueva

5

6 [a]Laboratory of Microbial Biofilms. Idab-CSIC-Universidad Pública de Navarra-Gobierno de Navarra

7 and Dpto Producción Agraria, Pamplona 31006, Spain.

8

9 [1] To whom correspondence should be addressed: Iñigo Lasa. Instituto de Agrobiotecnología,

10 Universidad Pública de Navarra-CSIC-Gobierno de Navarra, Pamplona-31006, Navarra, Spain. E-

11 mail: ilasa@unavarra.es; Fax: (34) 948 232191; Telephone: (34) 948168007

12

13

The precise understanding of the biology of a living cell requires the identification and quantification of the molecular components necessary to sustain life. One of such elements is RNA. Two independent high-throughput strategies are available to identify the entire collection of RNA molecules produced by a cell population, which is currently known as the transcriptome. One technique relies on microarray technology (tiling arrays) whereas the second one relies on sequencing the RNA pool (RNA-seq) (1). Both techniques offer the advantage that the identification of the RNA content is not biased by protein-based genome annotation. The application of these methods to the transcriptome analysis in bacteria has uncovered the existence of a large amount of RNA molecules that overlap at least in some portion with protein encoding RNA transcripts, generating perfect sense/antisense RNA duplexes (2-9). However, since transcriptome studies have been performed using microgram amounts of RNA purified from millions of bacterial cells instead of RNA purified from a single bacterium, the presence of overlapping sense/antisense RNAs from a genomic region does not necessarily mean that both sense/antisense transcripts are simultaneously present in the same bacteria. Hence, it might be possible that a subgroup in the bacterial population synthesized the sense transcript whereas another subgroup synthesized the antisense transcript, and consequently overlapping transcripts would never be together in the same cell. A report in PNAS by Lybecker et al provides clear evidences that both sense/antisense transcripts can be present simultaneously within the same bacterial cell (10). Using a monoclonal antibody that recognizes double stranded RNA molecules (dsRNA) irrespectively of the nucleotide sequence, the authors perform immunoprecipitation assays to pull down dsRNA molecules (IP-dsRNA) from a total RNA sample extracted from *Escherichia coli* and identified the purified dsRNA by RNA-seq.

Previous studies have identified examples of at least four different mechanisms to generate double stranded RNA duplexes in bacteria (2, 11): (i) the presence of "*bona fide*" non-coding antisense RNAs (asRNA); (ii) overlapping in the 5' region of mRNAs from contiguous genes that are transcribed in divergent directions; (iii) overlapping in the 3' regions of mRNAs from contiguous genes transcribed in convergent directions and finally, (iv) genes that being located in

1   the middle of an operon are transcribed in the opposite direction to the rest of the operon.

2   According to these mechanisms, the entire mRNA molecule seems to be susceptible to be

3   targeted by an overlapping transcript. However, with only a handful of transcriptomes available

4   so far, it is too early to establish whether overlapping transcription preferentially locates in a

5   specific region of the mRNA relative to the open reading frame. In this respect, Lybecker et al

6   find that the majority of IP-dsRNAs correspond to the 5 region of genes (50%), while only 0,5%

7   of the IP-dsRNAs correspond to overlapping transcripts that affect the 3' region (10). The most

8   common scenario is overlapping between long 5 UTRs of divergently transcribed genes,

9   followed by overlapping caused by asRNAs transcribed opposite to the 5 /Intergenic ends. This

10   is the first time that such strong bias of dsRNAs towards the 5' region has been described.

11   Whether differences in the size of overlapping regions might determine a more efficient

12   immunoprecipitation of dsRNA molecules at the 5' region cannot be excluded.

13   Pairing of overlapping transcripts provides double stranded RNA substrates that can be

14   digested by specific RNases. RNase III is a double-stranded RNA endoribonuclease, primarily

15   known for its roles in rRNA maturation, mRNA degradation, and sRNA processing (12). First

16   evidence that RNase III plays a critical role in the digestion of overlapping transcripts was

17   obtained in a recent study devoted to analyzing the transcriptome of the human pathogen

18   *Staphylococcus aureus* (7). This study found that RNase III digests overlapping transcripts

19   producing a collection of short RNA fragments (20 nucleotides on average), providing the first

20   evidence that both sense/antisense overlapping transcripts are present simultaneously in the

21   same cell. Other evidence supporting a genome-wide role of RNase III in processing RNA

22   molecules came from an independent study combining co-immunoprecipitation of a catalytically

23   inactive version of RNase III enzyme with deep RNA-seq (13). This study revealed that RNase III is

24   bound to many different antisense RNAs that cover 44% of annotated genes including non-

25   coding RNAs. In contrast, attempts to identify the collection of short RNA products generated by

26   RNase III mediated digestion of overlapping RNAs in *Salmonella* were unsuccessful, suggesting

27   that either overlapping transcripts might be processed by a different mechanism or that the

1   resulting short RNA molecules might be unstable in Gram-negative bacteria. The study of

2   Lybecker et al shows that indeed RNase III mediates the digestion of overlapping transcripts in *E.*

3   *coli*, indicating that this mechanism is well-conserved in bacteria. However, this study does not

4   clarify which are the end products of RNAse III activity because the antibody used is unable to

5   bind dsRNAs shorter than 40 base pairs. Intuitively, if RNase III digests overlapping transcripts,

6   one would expect that mutants in RNase III would accumulate higher levels of both

7   sense/antisense transcripts. In agreement with this hypothesis, Lybecker et al found that the

8   dsRNA regions are stable and more abundant than the single-stranded regions of the same

9   transcripts in the absence of an active RNase III. The simplest explanation is that dsRNA regions

10   remain protected while single-stranded regions are degraded by the activity of other RNases.

11   Overlapping transcription can affect the expression of its complementary gene at different

12   levels including transcription, mRNA stability or translation (14, 15). In this respect, the group of

13   P. Cossart has proposed a new paradigm of regulation based on overlapping transcription

14   termed "Excludon" (16, 17). The excludon concept describes the process by which the expression

15   of a long mRNA transcript results in the repression of the expression of the overlapping

16   transcript produced from the neighbor gene. The final consequence is that expression of both

17   overlapping transcripts is mutually exclusive. However, the exact mechanism underlying the

18   inhibitory effect of the overlapped transcripts in the excludon has not yet been determined.

19   Based on the observations that RNase III mediates the digestion of overlapping transcripts in *S.*

20   *aureus*, it was suggested that the selective degradation of the double-stranded RNA that results

21   from hybridization of overlapping sense and antisense transcripts could be a likely mechanisms

22   to explain the excludon paradigm (17). The study of Lybecker strongly supports this idea. An

23   interesting question regarding the RNase III processing of overlapping transcripts is whether the

24   digestion occurs before or after mRNA translation. In general, it is assumed that the transcription

25   and translation processes are coupled in bacteria. In this scenario, RNase III would be processing

26   RNA molecules that had already been translated, and pairing between overlapping transcripts

27   would represent another mechanism of RNA decay. Alternatively, if the digestion of RNAs by

RNase III occurs before translation, this mechanism would provide an additional level of post-transcriptional regulation to adjust mRNA levels and/or to remove any transcript produced due to leaky transcription initiation. The former point is very important because uncontrolled transcription might be toxic if all mRNAs were translated at the same time. Depending on the expression levels of each overlapping transcript Lybecker et al identified two classes of dsRNAs. In class I, both overlapping RNAs exhibit different expression levels whereas in class II both transcripts have similar expression levels. Complementary proteomic studies will be necessary to determine whether the transcripts of class I that are produced at lower levels are indeed translated to proteins. On the other hand, it is important to highlight that, coinciding with previous observations by Lioliou et al (13), a significant number of antisense RNAs transcribe opposite to non-coding regulatory RNAs, indicating that the RNAse III mediated digestion of overlapping transcripts may impact the functionality of the RNA molecules regardless of the protein translation process.

The study of Lybecker et al provides a novel method to study the genome-wide process of overlapping transcription in bacteria. This technology has provided a tool to demonstrate that sense and antisense transcripts exist simultaneously in the cytoplasm of *E. coli*. Identical conclusions were previously obtained in *S. aureus* using a completely different strategy, indicating that degradation of overlapping transcripts by RNase III is a highly conserved process in bacteria (7). Of course, many intriguing questions about the process remain. What are the end products of the RNase III digestion process in *E. coli*? Are these end products functional molecules with a specific role in gene regulation? When does the pairing between overlapping transcripts occur, before or after translation? What is the kinetics of the RNase III processing reaction? Are there specific proteins governing the pairing between overlapping transcripts? Which of the phenotypes associated to RNase III deficiency are due to the accumulation and translation of sense and antisense transcripts in the same cell? The initial skepticism about the biological relevance of the genome-wide overlapping transcription has been followed by the discovery that bacteria have a simple and efficient mechanism to remove the double stranded

1 **sense/antisense pairs. Because this process can alter the levels of functional RNAs, the time has**

2 **come to include overlapping transcription as another player of bacterial gene regulation.**

3

4

5 **Figure legend.** Schematic description of the dsRNA immunoprecipitation assays (IP-dsRNA) from a

6 total RNA sample extracted from wildtype and RNase III mutant strains. RNase III mediated

7 digestion of overlapping transcripts reduces the amount of dsRNA in the bacterial transcriptome.

8

9 **References**

10
11 1.   Sorek R, Cossart P (2010) Prokaryotic transcriptomics: a new view on
12   regulation, physiology and pathogenicity. *Nature Reviews Genetics* 11:9–16.

13 2.   Toledo-Arana A et al. (2009) The *Listeria* transcriptional landscape from
14   saprophytism to virulence. *Nature* 459:950–956.

15 3.   Dornenburg JE, Devita AM, Palumbo MJ, Wade JT (2010) Widespread
16   antisense transcription in *Escherichia coli*. *mBio* 1.

17 4.   Sharma CM et al. (2010) The primary transcriptome of the major human
18   pathogen *Helicobacter pylori*. *Nature* 464:250–255.

19 5.   Georg J et al. (2009) Evidence for a major role of antisense RNAs in
20   cyanobacterial gene regulation. *Molecular Systems Biology* 5:305.

21 6.   Güell M et al. (2009) Transcriptome complexity in a genome-reduced
22   bacterium. *Science* 326:1268–1271.

23 7.   Lasa I et al. (2011) Genome-wide antisense transcription drives mRNA
24   processing in bacteria. *Proceedings of the National Academy of Sciences*
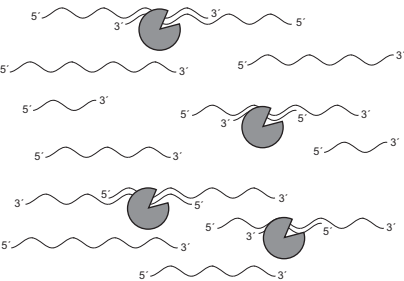25   108:20172–20177.

26 8.   Liu JM et al. (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by
27   direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res*
28   37:e46.

29 9.   Jäger D et al. (2009) Deep sequencing analysis of the *Methanosarcina mazei*
30   Go1 transcriptome in response to nitrogen availability. *Proc Natl Acad Sci*
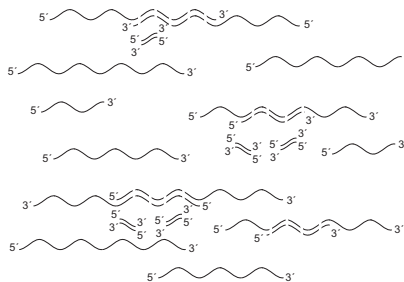31   *USA*.

1   10.   Lybecker M, Zimmermann B, Bilusic I, Tukhtubaeva N, Shroeder R (2013) The
2   double-stranded transcriptome of *E. coli*. 1–6.

3   11.   Lasa I, Toledo-Arana A, Gingeras TR (2012) An effort to make sense of
4   antisense transcription in bacteria. *RNA biology* 9:1039–1044.

5   12.   Arraiano CM et al. (2010) The critical role of RNA processing and degradation
6   in the control of gene expression. *FEMS Microbiology Reviews* 34:883–923.

7   13.   Lioliou E et al. (2012) Global Regulatory Functions of the *Staphylococcus
8   aureus* Endoribonuclease III in Gene Expression. *PLoS Genetics* 8:e1002782.

9   14.   Thomason MK, giselaa S (2010) Bacterial antisense RNAs: how many are
10   there, and what are they doing? *Annual review of genetics* 44:167–188.

11   15.   Georg J, Hess WR (2011) *cis*-antisense RNA, another level of gene regulation
12   in bacteria. *Microbiol Mol Biol Rev* 75:286–300.

13   16.   Wurtzel O et al. (2012) Comparative transcriptomics of pathogenic and non-
14   pathogenic *Listeria* species. *Molecular Systems Biology* 8:583.

15   17.   Sesto N, Wurtzel O, Archambaud C, Sorek R, Cossart P (2013) The excludon:
16   A new concept in bacterial antisense RNA-mediated gene regulation. *Nat
17   Rev Micro* 11:75–82.
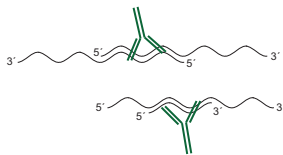
18

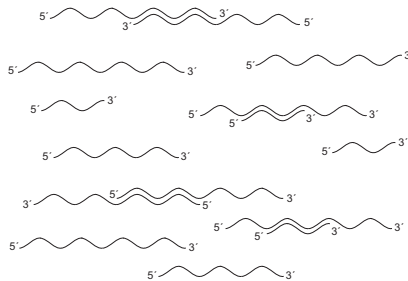Bacterial transcriptome