Navigation through the twists and turns of RNA sequencing technologies: Application to bacterial regulatory RNAs

Emma Desgranges, Isabelle Caldelari Baumberger, Stefano Marzi, David Lalaouna

Navigation through the twists and turns of RNA sequencing technologies: Application to bacterial regulatory RNAs

Emma Desgranges, Isabelle Caldelari, Stefano Marzi and David Lalaouna*

Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR9002, Strasbourg, France

* To whom correspondence should be addressed. E-mail: d.lalaouna@ibmc-cnrs.unistra.fr; Tel: +33388417051

ABSTRACT

Discovered in the 1980s, small regulatory RNAs (sRNAs) are now considered key actors in virtually all aspects of bacterial physiology and virulence. Together with transcriptional and translational regulatory proteins, they integrate and often are hubs of complex regulatory networks, responsible for bacterial response/adaptation to various perceived stimuli. The recent development of powerful RNA sequencing technologies has facilitated the identification and characterization of sRNAs (length, structure and expression conditions) and their RNA targets in several bacteria. Nevertheless, it could be very difficult for non-experts to understand the advantages and drawbacks related to each offered option and, consequently, to make an informed choice. Therefore, the main goal of this review is to provide a guide to navigate through the twists and turns of high-throughput RNA sequencing technologies, with a specific focus on those applied to the study of sRNAs.

KEYWORDS

small regulatory RNA; RNA sequencing; post-transcriptional regulation; targetome; regulatory network; bacteria

INTRODUCTION

The continuous advances in the characterization of bacterial small regulatory RNAs (sRNAs) have allowed the appreciation of their diversity regarding their origins, sizes and functions [1, 2]. Usually, the transcription of sRNAs relies on two-component systems and/or transcriptional factors responding directly to environmental stimuli or metabolites [2]. This heterogeneous group of RNAs includes not only *bone fide* sRNAs, acting in *trans* and having their own promoter and transcriptional terminator, but also RNA fragments released from precursor RNAs. Typically, these fragments are the result of an RNase-dependent cleavage (e.g. RNase E, RNase III) of 3' untranslated region (UTR) of mRNAs [3], 3' external transcribed spacer of tRNA transcripts [4] or sRNAs, which are processed to give mature forms. For example, ArcZ sRNA is transcribed as a 120 nt-long fragment in *Escherichia coli* and then processed into two shorter forms, a low abundant (88 nts) and a stable (55 nts) fragments, sharing the same 3' end [5]. The genesis of an sRNA (proper transcription or maturation) can be easily discriminated by analyzing their distinctive 5' ends. Primary RNAs, which are produced from their own promoter, have a 5' triphosphorylated (5'PPP) end, while secondary RNAs which are processed from a precursor transcript, have a 5' monophosphorylated (5'P) end. Hence, 5' ends are characterized as transcription start sites (TSS) and processing start sites (PSS), respectively.

sRNAs do not only differ in their origins but also in their lengths. They are commonly described as ranging from 50 to 500 nts. However, recent findings revealed that they are even more heterogeneous in size, from dozens to over a thousand nucleotides. Indeed, very small RNA fragments (such as tRNA-derived fragments [6]) and large transcripts (e.g. 1,116 nt-long RsaC sRNA in *Staphylococcus aureus* [7]) have been described as trans-acting regulators.

sRNAs act mainly by forming non-contiguous pairings with their target mRNAs, thus regulating their translation and/or stability. Chaperone proteins like Hfq or ProQ are often involved in sRNAs stability and/or functions [8, 9]. Remarkably, an sRNA can regulate many mRNA targets, constituting its targetome, and one target can be regulated by numerous sRNAs. As a result, sRNAs post-transcriptionally control the expression of their multiple mRNA targets involved in all kinds of cellular processes and can even regulate their own transcription factors via feedback loops. They are often non-coding, meaning that they do not bear ribosome binding signals and in frame start and stop codons, the key features of coding sequences. Yet few are named bifunctional or dual-function sRNAs, as they encode peptides like SgrS in *E. coli* and RNAIII in *S. aureus* [10], coding for the SgrT peptide and the Hld toxin, respectively. In addition, a few sRNAs buffer other sRNA transcripts [4, 11] or cellular proteins [12, 13]. In a nutshell, sRNAs integrate various internal and/or external signals and are at the center of complex regulatory networks, impacting every aspect of bacterial physiology/virulence.

This review will provide an overview of cutting-edge technologies/approaches for the complete characterization of bacterial sRNAs, their whole targetome and their cellular functions. Due to the profusion of available algorithms/methods, we provide here a non-comprehensive list, and we sincerely apologize to authors of uncited works.

1. Identification of sRNA pool in a bacterial genome

Historically, sRNAs have been mainly identified by computational searches (potential promoters and rho-independent terminators in intergenic region), comparative genomics (sequence conservation) or co-purification with RNA-binding proteins such as Hfq and CsrA/RsmA [14, 15]. However, the introduction of tiling microarrays and RNA sequencing (RNA-seq)-based approaches has revolutionized our understanding of bacterial transcriptomes, including sRNAs. With the development of next-generation sequencing (NGS) technologies, the cost associated to transcriptomic analysis has decreased at a dramatic rate, making RNA-seq more affordable. As a consequence, we have recently observed an explosion of transcriptomics information. Indeed, the study of sRNAs is no longer restricted to model organisms such as *E. coli* or *Salmonella* Typhimurium. The transcriptome of many pathogenic strains such as *S. aureus* [16], *Helicobacter pylori* [17], *Mycobacterium tuberculosis* [18], *Streptococcus pneumonia* [19], *Clostridium difficile* [20], *Pseudomonas aeruginosa* [21], but also of too often neglected environmental strains [22-24], has led to the discovery of many non-coding transcripts with potential regulatory functions. Remarkably, the improvement in sensitivity of sequencing technologies has revealed that bacterial sRNAs are not only restricted to the cytoplasm but are also packaged in secreted membrane vesicles to suppress host immune defenses [6, 25]. Indeed, *P. aeruginosa* produces outer membrane vesicles containing sRNAs, including a tRNA-derived fragment, which is released in the host cell and reduces IL-8 secretion [6]. Similarly, the intracellular pathogen *Listeria monocytogenes* secretes both “naked” RNAs and intravesicular RNAs [25].

RNA-seq data mining and analysis is still arduous and time consuming. Algorithms and tools such as Rockhopper [26], ANNOgesic [27], sRNA-Detect [28], and APERO [29] were developed to facilitate the genome-wide identification of sRNAs which are inferred from RNA sequencing data. Leonard *et al.* [29], compared commonly used algorithms using *E. coli* and *S.* Typhimurium RNA-seq data, and concluded that the number of identified sRNAs is extremely variable, from hundreds to thousands. This is typically due to bias of respective detection methods. Indeed, most of them are using read coverage distribution to define sRNAs, searching either for a uniform coverage or setting a coverage threshold. Even if these approaches are quite efficient for eukaryotic sRNAs (e.g. miRNAs and siRNAs), bacterial

sRNAs are longer and often exhibit a non-homogeneous coverage due to their strong secondary structure, their processing by endoribonucleases (e.g. RNase E and RNase III) and exoribonucleases (e.g. PNPase) or the presence of repetitive elements. This ultimately leads to an overestimation of their number and an incorrect deduction of their extremities. The type of library preparation (size-selection, fragmentation, adaptor ligation or random priming, single- or paired-end sequencing) can also impact the outcome. Among the above-mentioned analysis algorithms, a main advantage of APERO (Analysis of Paired-End RNA-Seq Output) is to directly use individually sequenced non-fragmented transcripts (paired-end). Hence, APERO circumvents the lack of coverage homogeneity, and infers more accurately the 5' and 3' extremities of sRNAs. The size-selection of non-fragmented transcripts in the sRNA library preparation (50 to 200 nts) also helps to identify RNA fragments deriving from processed precursors, which are otherwise often lost in the mass of fragmented reads corresponding to the respective primary transcripts in a global transcriptomic approach. However, a main drawback of this method is the size-selection, with the possible exclusion of very small and large regulatory RNAs. Tested experimental conditions can also limit sRNAs identification, as specific stimuli trigger their production. To circumvent the limitation, RNA extracted in different growth conditions could be pooled before the analysis.

Despite the improvements made in sRNA prediction tools based on RNA-seq data, the proper identification of sRNAs and their boundaries still require dedicated experimental protocols and *in vivo* validations.

2. Characterization of sRNAs at a single molecule or a global scale

Full sRNAs characterization is a prerequisite for the experimental design to unveil their targetome and functions (section 3). For example, a careful determination of the 5' and 3' end of sRNAs (and their mRNA targets) is required for cloning and mutagenesis purposes and, consequently, for the characterization of sRNA-dependent regulatory mechanisms. This is also important for the study of sRNAs genesis and expression.

In the following paragraphs, we will describe suitable methods allowing the accurate determination of sRNA extremities and their secondary structure, at a single molecule or a global level (Figure 1). To note, most mentioned methods are not specific to sRNAs and could be applied to any RNA transcript.

2.1. Characterization of their 5' extremity

One of the biggest challenges in the sRNAs characterization is to determine their extremities. For this purpose, specific protocols have been developed to identify 5' ends and

discriminate between transcription and processing start sites, while other methods allow the direct sequencing of 3' ends (section 2.2). Interestingly, both extremities can also be simultaneously determined (section 2.3).

Several approaches are available for the determination of the 5' end of individual transcripts, such as 5' RACE (Rapid amplification of cDNA ends) and primer extension, which are here concisely described. There are many variants of the 5' RACE method. Commercial kits are also available, such as 5' RACE System for Rapid Amplification of cDNA Ends (Invitrogen). Here, we describe an efficient and user-friendly protocol routinely used in many laboratories [30]. This method is inexpensive and relatively easy to implement. A complementary DNA (cDNA) is obtained by reverse transcription (RT) using a sequence-specific oligonucleotide from total RNA. The design of this primer is a decisive step. It must be highly selective for your RNA of interest and located in the vicinity of the +1 (50-150 nucleotides downstream), which requires prior knowledge concerning the putative 5' end. If not, different oligos can be used. An adaptor of known sequence, acting as an anchor sequence, is then linked at the 3' end of the single stranded cDNA by the T4 RNA ligase. A complementary oligonucleotide is used to initiate the synthesis of the second strand. Finally, the double-stranded cDNA is amplified by nested PCR and finally inserted into a cloning vector. The nested PCR corresponds to two successive rounds of PCR using two distinct and specific primers along with the oligonucleotide complementary to the 3' adaptor. The first amplicon is used as template for the second PCR primed by an internal oligonucleotide. This approach highly reduces non-specific amplifications. After sequencing, the +1 of your favorite RNA is identified as the first nucleotide downstream the adaptor sequence. As reverse transcriptases are subject to premature termination, the 5′ end could be misattributed. Hence, the ligation of an RNA adaptor directly to total RNA 5′ ends using T4 RNA ligase prior to RT could be another strategy for reducing non-specific amplifications [31]. However, in return, the reaction will mainly depend on ligation efficiency and RNA abundance.

Primer extension assays are also commonly used to determine the +1 of specific transcripts by RT run-off signal analysis. A detailed protocol is provided by Carey *et al.* [32]. Unlike 5' RACE approach, primer extension assays are generally performed by using a P^{32} radio-labeled DNA oligonucleotide to prime the RT reaction. The 5' terminus of a sequence-specific DNA primer is thus radiolabeled with γ-^{32}P ATP by the T4 polynucleotide kinase (PNK). In this way, the chosen RNA is reverse transcribed from total RNA, leading to the synthesis of a radioactive single-stranded cDNA. The sample is then loaded on a high-resolution polyacrylamide gel next to sequencing ladders (G, A, T and C) obtained by Sanger method. Hence, the +1 of the starting transcript is identified without the need for cloning or sequencing. The method can be tricky if the RNA of interest is weakly expressed (due to the absence of PCR amplification), or if it is highly structured (detection of several bands due to premature

terminations of RT). A safer and time-saving alternative is to use fluorescently labeled primers and DNA sequencing apparatus [33].

In the last decade, genome-wide RNA sequencing approaches such as dRNA-seq [34], Cappable-seq [35] and TagRNA-seq [36] have been developed to identify the 5' extremity of all transcripts. They also allow investigators to distinguish between primary and processed RNAs and, by extension, between transcription and post-transcriptional maturation at a global scale. In parallel, an *in silico* prediction of putative bacterial promoter elements (-10 and -35 boxes) upstream the identified TSS can be performed.

The first developed method, called differential RNA-seq (dRNA-seq) [34], identifies TSS at single-base resolution, and discriminates the different types of 5' end by comparing a control library, representing all transcripts, with a library obtained after specific degradation of processed transcripts. This depletion is achieved using a Terminator 5'-phosphate-dependent exonuclease (TEX, Lucigen Corporation) treatment. In both cases, all 5' triphophates are converted to 5' monophosphates by the tobacco acid pyrophosphatase (TAP; Epicentre) before cDNA synthesis to allow the ligation of an RNA adapter. In addition, RNAs are poly(A) tailed by the poly(A) polymerase, to synthetize the first-strand cDNA using a complementary oligo(T) primer. dRNA-seq method has been widely used to characterize sRNAs in bacteria such as *Helicobacter pylori* [37] and *Neisseria meningitidis* [38], and especially those generated by RNase-dependent processing like 3'UTR-derived sRNAs. The major limitation of this approach is that modified RNAs (5' capped RNAs, see below) or RNAs with a strong secondary structure such as tRNAs, are insensitive to the exonuclease attacks and indistinguishable from 5' triphosphorylated transcripts.

Cappable-seq [35] is using an opposite strategy. Instead of degrading 5' monophosphorylated RNAs, Cappable-seq captures biotinylated 5'triphophorylated RNAs. Total RNA is first treated by the vaccinia capping enzyme, resulting in the capping of 5'PPP ends with a biotinylated GTP. After enrichment using streptavidin-coated magnetic beads, capped-5' extremities are converted to 5' monophosphorylated ends by RppH treatment. As a control, the streptavidin-dependent enrichment step is skipped. For cDNA libraries preparation, an adaptor is ligated to both 5' and 3' extremities. Remarkably, Cappable-seq was applied to the mouse gut microbiota [35] and the *Wolbacchia* bacterial endosymbiont of nematodes [39], enabling TSS determination in complex environments. A main advantage is that this fishing step implies no need for rRNAs depletion. Still, a contamination with processed RNAs is observed (at least 3-5% of enriched RNAs).

TagRNA-seq method [36] is based on the ligation of different tags to PSS and TSS, enabling to discriminate both types of 5' extremities in a single experiment. This represents a main advantage when compared to other technologies which eliminated 5'

monophosphorylated extremities and then compare two distinct libraries, a source of bias. 5' monophosphorylated RNAs are first labelled with a specific tag. Subsequently, 5' triphosphates are converted to 5' monophosphates (TAP treatment) and newly obtained 5'P end are labelled with a different tag. In theory, TagRNA-seq clearly distinguishes TSS and PSS, but in practice it is not so straightforward due to several bias (e.g. aspecific ligation, spontaneous cleavage). Consequently, a mixture of TSS and PSS is generally mapped at the same position. Additional bioinformatic analyses are then required.

Remarkably, a cap-like structure has been detected at the 5' end of some sRNAs, which is notably involved in their stability. According to Bird and colleagues [40], the non-canonical initiating nucleotide (NCIN) is incorporated into RNA during transcription initiation by RNA polymerase. The percentage of capped sRNAs seems to be rather limited. Indeed, a few *E. coli* sRNAs are known to be modified in their 5' with nicotinamide adenine dinucleotide (NAD+) [41]. In addition, this phenomenon is restricted to a small proportion of the sRNA population (e.g. ~13% of RNAI sRNA). This discovery was made possible by the development of chemo-enzymatic capture and next-generation sequencing NAD captureSeq method [41]. Essentially, the NAD+ group is transglycosylated and then biotinylated to enable the streptavidin-dependent purification of NAD-modified 5' end. Another approach, called CapZyme-Seq [42], proposes to extend the study to other nucleoside-containing metabolites, such as NADH, 5′-desphospho coenzyme A (dpCoA) and flavin adenine dinucleotide (FAD), which could potentially function as NCIN capping [40]. To globally detect and quantify NCIN-capped RNAs, CapZyme-Seq method compares two datasets generated by selective enzymatic processing: NCIN-capped 5′ ends are treated by the RNA-decapping enzyme NudC and uncapped 5′ ends by the RNA 5′ polyphosphatase (Rpp) to obtain 5' monophosphorylated RNAs. Hence, Vvedenskaya *et al.* notably revealed three major capped-sRNAs, SibE, SibD and OxyS (13-22%) in *E. coli*. A main advantage of CapZyme-Seq is to concurrently identify the sRNA 5' end at the single-nucleotide resolution, but 5'PPP and 5'P are not discriminated. While CapZyme-Seq allows the detection of RNAs carrying other NCIN-caps than NAD+, the nature of the non-canonical initiating nucleotide remains unknown. Indeed, NudC recognizes at least four distinct NCIN-caps: NAD+, NADH, dpCoA, and FAD. For this purpose, the use of additional and specific processing enzymes should be considered. Another limitation is that processing efficiency could be altered by the presence of strong secondary structures.

2.2. Characterization of their 3' extremity

The vast majority of sRNA possesses a Rho-independent terminator (or intrinsic terminator) [43], composed of a GC-rich palindrome sequence followed by a stretch of U residues (≈6-8 residues). This structural element has the obvious role of marking the 3'end of

an sRNA during transcription, but also of stabilizing it, notably being a recognition site of Hfq [44]. A non-negligible proportion of termination is mediated by the ATP-dependent DNA translocase Rho, which likely interacts with the RNA polymerase to induce its release [45]. To our knowledge, no sRNA has been yet described to possess a Rho-dependent terminator. The sequence recognized by Rho (Rho utilization site, rut) is still poorly defined (≈60-80 nts, C/G>1, C residues every 11-13 nts), making it difficult to accurately identify Rho-dependent termination. Remarkably, the 3' end of an sRNA can differ notably due to transcriptional read-through (e.g. RsaA sRNA possesses two distinct terminators in *S. aureus* [46]).

Intrinsic terminator can be spotted by various bioinformatic tools such as TransTermHP [47], RNIE [48] or iTerm-PseKNC [49]. Very recently, RhoTermPredict was developed to predict Rho-dependent termination, notably by searching for putative rut sites followed by a pause site for the RNA polymerase [50]. Because of a significant rate of false positives, *in silico* predicted terminators must be experimentally validated.

A high-throughput technology called Term-seq has been developed to identify the 3' extremity of every RNA fragment at a single-base resolution [51]. Briefly, an adaptor is first added to each 3' end. RNAs are sheared and then reverse transcribed using a complementary oligonucleotide. Here, another adapter is ligated to the 3' end of the newly synthetized single-strand cDNA to produce a double-stranded cDNA. Thus, the obtained cDNA library is sequenced. Term-seq has notably allowed the discovery of several transcriptional riboswitches. These regulatory elements are found in the 5'UTRs of several mRNAs and adopt a distinct secondary structure in presence of a specific ligand, affecting the transcription or the translation of controlled gene(s).

2.3. Simultaneous determination of 5′ and 3′ ends

As it could be laborious to separately determine 5' and 3' ends, several techniques have been developed to concurrently characterize both boundaries of RNAs.

The main advantage of the 5'3' RACE (also named circular RACE) is to determine both 5' and 3' ends which does not rely on the ligation of adapters. First, the 5' triphosphorylated end is converted to 5' monophosphate by TAP or RppH treatment. The RNA of interest is then circularized by ligation of its 5′P and 3'OH extremities, which is catalyzed by RNA Ligase [52]. To note, polyethylene glycol (PEG) can be used to improve the ligation efficiency. The reverse transcription is performed using a sequence-specific primer. Then, the site of ligation is amplified by nested PCR, and is finally sequenced.

Library preparation by previously described methods generally require the fragmentation of transcripts, which prevent the determination of both 5' and 3' ends. Remarkably, the high-throughput approach called simultaneous 5′ and 3′ end sequencing (SEnd-seq) allows the mapping of both extremities at a global scale and without size-selection [53]. First, a 3' adaptor is added to total RNA using T4 RNA Ligase 1. The first strand cDNA is obtained by RT using a biotinylated primer complementary to the 3' adaptor. After 5' phosphorylation by T4 PNK, cDNAs are circularized by TS2126 RNA Ligase I, favoring intramolecular ligation. Circularized cDNAs are fragmented and the second strand is obtained using the Second Strand cDNA Synthesis Kit (NEB). Biotin-labeled DNA fragments, containing the 5'3' junction, are purified by streptavidin beads and sequenced. Interestingly, SEnd-seq also allows to distinguish between 5'PPP and 5'P extremities by a specific enrichment of primary or processed transcripts. Performed in *E. coli*, SEnd-seq method identified sRNAs of various lengths such as CyaR (87 nts), GcvB (204 nts), CsrB (365 nts). Oddly, SEnd-seq obtained results that differ from previously characterized lengths of GcvB (205 nts) and CsrB (369 nts). As putative technical artefacts cannot be ruled out, these results must be individually validated by some of the previously mentioned methods before drawing any conclusions.

2.4. Determination of their secondary structure

Intramolecular interactions (formation of hydrogen bonds between complementary nucleotides, usually A-U, G-C and G-U) generate stem-loops and more complex structures (e.g. pseudoknots), which are crucial for sRNAs stability, protein recognition and constrain pairing with their targets [54].

Software using individual (e.g. UNAFold [55] and RNAFold [56]) or comparative structure predictions (e.g. RNAalifold [57]) predict RNA secondary structure based on consensus structures, co-variation and minimum free energy. Interestingly, some *in silico* tools integrate information provided by structure probing assays such as SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension; e.g. RNAstructure [58]). A more comprehensive list of available algorithms is provided by Wright *et al.* (2018) [59].

In vitro RNA probing techniques using chemicals (CMCT, DMS, Pb(II), BzCN) or RNases (A, T1, T2, V1) are performed with any kinds of RNA molecules from any organisms (for review [60]). Direct radioactive labeling of smaller RNAs (<300 nucleotides) followed by loading of cleaved RNA (from RNases action or from chemical treatment of modified bases) on a denaturing polyacrylamide gel is effective, while probing of longer RNAs requires primer extension by reverse transcriptase.

Several high-throughput methods have appeared in the last years to study RNA structure at a large scale. PARS (Parallel Analysis of RNA Structures) has been implemented in bacteria to determine the secondary structure of *in vitro* refolded total RNA, using two distinct cocktails of enzymes cleaving either single stranded or double-stranded RNAs [61]. Digested RNA fragments are phosphorylated with T4 PNK prior to library preparation and deep sequencing. In *Yersinia pseudotuberculosis*, Righetti *et al.* identified multiple RNA riboswitches, called thermometers, controlling gene expression in a temperature-dependent manner [61]. Remarkably, PARS-derived structures of abundant sRNAs such as CsrB and CsrC were also determined. However, a major drawback is that RNA structures are formed *in vitro* after denaturation of total RNA, which could generate artefactual interactions (both intra- and intermolecular). The cellular abundance of sRNAs is also a limiting factor. The ultimate goal is obviously to explore the secondary structure of RNAs in a cellular context. For this purpose, several *in vivo* approaches have been recently developed and are reviewed by Ignatova and Narberhaus [62]. Among them, in-cell SHAPE-seq uses the ability of 1-methyl-7-nitroisatoic anhydride (1M7) to cross membranes and penetrate living cells [63]. 1M7 is a SHAPE reagent which modifies the 2'-hydroxyl group of flexible (unpaired) RNA nucleotides. Modified 2'OH groups, and by extent, reactive nucleotides, are then detected by primer extension using their ability to arrest reverse transcription. Here, Watters and colleagues used specific RT primers to target RNAs of interest such as the *btuB* mRNA riboswitch domain in *E. coli*, which facilitate data analysis. However, to our knowledge, no suitable approach exists to determine the *in vivo* secondary structure of sRNAs at a global scale. Indeed, the presence of multiple interaction sites (intermolecular interactions with RNA binding proteins and mRNAs) could lead to a misinterpretation of RNA flexibility and, thus drastically complexify data analysis. Therefore, the *in vivo* characterization of sRNAs structure still represents a big challenge in the field.

3. Characterization of their targetome

Several features make the characterization of sRNAs targetome difficult: (1) *trans*-acting sRNAs often interact with multiple target mRNAs (e.g. >30 targets for GcvB sRNA in *E. coli* [64]) using limited and non-contiguous sequence complementarities; (2) a single sRNA can bear multiple pairing sites, also named "seed" sequences; (3) RNA chaperone proteins such as Hfq [9] and ProQ [8] are often described as promoting the pairing between an sRNA and its partners especially in *Enterobacteriaceae*. During the last decade, several algorithms and experimental approaches have been developed to tackle this major problem.

The pairing site between an sRNA and its targets can be predicted by *in silico* tools such as TargetRNA2 [65], sTarPicker [66], IntaRNA [67] and CopraRNA [68]. These tools are

widely compared in the following reviews [59, 69, 70]. According to Pain *et al.* (2015) [69], CopraRNA outperforms other algorithms, due to the integration of both the accessibility of the interaction sites and the evolutionary conservation of sRNA:mRNA interaction. IntaRNA is an attractive alternative, notably if sequence comparison by different organisms is not possible [69, 70]. Remarkably, a recently developed software pipeline, called SPOT for "sRNA target prediction organizing tool" [71], regroups all previously cited algorithms and also integrates experimental data (from MAPS or RIL-seq, see below) to improve targets prediction. However, the false-positive rate remains high. Moreover, a significant part of the RNA target is overlooked, as most algorithms focus on a specific window around the start codon (-200 to +100; default parameters). They also focus only on coding RNAs and completely ignore sRNAs, tRNAs or even rRNAs as putative targets. However, such interactions have been recently described [4, 72].

During the last decade, many experimental approaches have been developed to deeply characterize the targetome of sRNAs. The major techniques are based either on the purification of a tagged sRNA with its partners or on the pull-down of all RNAs interacting with an RNA binding protein (RBP). They have been recently reviewed by Saliba *et al.* (2017) [73].

3.1. Deep analysis of a specific sRNA targetome

In numerous organisms, no sRNA-associated protein (e.g. Hfq or ProQ homologs) has been identified. Hence, an elegant strategy is to use an sRNA as bait to co-purify all interacting RNAs. MAPS (MS2-affinity purification coupled with RNA sequencing) technology was performed in several organisms like *E. coli* [4, 64, 74, 75], *Salmonella* Typhimurium [76] and *S. aureus* [77, 78]. Detailed protocols are available [79, 80]. Briefly, the sRNA of interest is fused to an MS2 aptamer and synthetized *in vivo* (Figure 2). Interacting RNAs are then co-purified using the strong and specific affinity of MS2-MBP (maltose binding protein) fusion protein (immobilized on an amylose column) for MS2 aptamer. Enriched RNAs are finally eluted by addition of maltose, a binding competitor. MAPS technology deeply characterizes and zooms in on a specific targetome. The co-purification step enables the detection of even low abundant targets. Major limitations are (1) the addition of the MS2 tag could affect function, stability and/or folding of sRNA. Consequently, all MS2-sRNA constructs must be scrupulously validated; (2) some targets might be missed by the use of improper growth conditions; (3) MS2-sRNA are often overexpressed when compared to endogenous expression; (4) indirect co-purification could occur via interactions with RNA-binding proteins or other RNAs, then an experimental validation is still necessary.

An alternative approach called Hybrid-trap-seq [81] is based on the co-purification of interacting RNAs with a 3'-biotinylated RNA *in vitro* (Figure 2). The sRNA of interest is *in vitro* transcribed and 3' end biotinylated. Total RNA is extracted in various experimental conditions and mixed with the synthetic sRNA bait. *In vitro* formed complexes are then isolated using streptavidin magnetic beads. Notably, Hybrid-trap-seq was performed in *S. aureus* to identify several targets of RsaE sRNA. Major limitations of this approach are: (1) the use of a synthetic RNA. Improper RNA folding and the absence of post-transcriptional modifications could have a detrimental impact on the results; (2) RNA:RNA complexes are formed *in vitro* after denaturation. RNA folding and pairing conditions (RNAs renatured altogether at room temperature) could generate some interactions with regions which are paired *in vivo*. (3) RNA chaperone proteins, which could be involved in RNA folding and RNA:RNA interaction, are not present during *in vitro* annealing.

3.2. sRNA:RNA interactions associated with an RBP

RIL-seq (RNA interaction by ligation and sequencing) [82-84] and CLASH (UV-crosslinking, ligation and sequencing of hybrids) [85] aim to define the global RNA-RNA interactome associated with a specific RBP in bacteria. Both methods share major similarities. They rely on UV crosslinking between an RBP (e.g. ProQ, Hfq or RNase E) and its RNA partners, followed by a ligation between proximate RNAs (Figure 2). The RNA:RNA chimeras are sequenced and identified by mapping reads on the reference genome. By comparing the first two studies (Hfq/RIL-seq [82] and RNaseE/CLASH [85]), Hör and Vogel (2017) [86]) concluded that even if two different bait proteins were used, the number of sRNA:mRNA pairs is very similar, suggesting that Hfq and RNase E act together on these specific hybrids. This is consistent with previous studies showing that Hfq recruits RNase E to induce sRNA-dependent mRNA decay [74, 87]. Major limitations of both methods are (1) they require tagging of the RBP of interest, which can prevent their activity; (2) the identification of sRNA:RNA interactions is restricted to sRNAs that function with the RBP of interest, in specific growth conditions; (3) the number of interactions might be under-estimated as crosslinking and ligation steps are not fully efficient; (4) the binding of RBP to RNA duplexes might be transient.

3.3. Unravelling all RNA:RNA interactions

Various approaches are now developed to eliminate the co-purification step (either with an interacting protein or RNA) and provide atlases of every RNA:RNA interaction.

Hi-GRIL-seq (High-throughput global small non-coding RNA target identification by ligation and sequencing) is based on the expression under an inducible promoter of the T4 RNA ligase, which covalently ligates the 3'-OH of an RNA to the 5'monophosphate of another

one (Figure 2) [88, 89]. After removal of rRNAs, the ligated mixture is sequenced. In contrast to RIL-seq, CLASH or MAPS, the method does not require affinity purification steps, avoiding the disassembly of duplexes. However, two major drawbacks of the experiment are the overproduction of the T4 RNA ligase *in vivo*, which could lead to unspecific RNA pairs and the absence of trimming of 5' triphosphate RNAs before ligation. Indeed, without a trimming step, the proximity ligation rate might be variable, mostly depending on the distance between the pairing site and RNA extremities. Here, only 0.26-0.29 % of reads are chimeric fragments. Among them, the regulon of the sRNA PrrF1 of *Pseudomonas aeruginosa* was extended and new RNA:RNA interactions revealed [88].

Liu and co-workers [90] developed in *E. coli* a modified version of the CLASH method, which does not require a co-purification step using a specific tagged-protein. Therefore, this approach has the potential to detect every RNA-RNA interaction and, consequently, to unravel the whole interactome. Basically, cells are treated with 4'-aminomethyl trioxsalen (AMT) and then exposed to UV (365 nm), which leads to the formation of adducts between opposing uridine residues (Figure 2). AMT specifically crosslinks RNA:RNA interactions (both intra- and intermolecular), and not RNA:protein or protein:protein interactions such as formaldehyde. After extraction, total RNA is trimmed with RNase T1, that specifically degrades single-stranded RNA at G residues. Residual single-stranded RNAs are eliminated by hybridization of 20-mer oligonucleotides and RNase H treatment. RNA duplexes are purified on denaturing gel according to their size (40-100 nts). Free RNA extremities are proximity ligated by T4 RNA Ligase 1, before crosslink reversal (UV irradiation at 254 nm). Chimeric fragments are finally submitted to RNA sequencing. To assess bias of the modified CLASH method, several controls are used (e.g. without AMT and/or T4 RNA ligase). After bioinformatics analysis, only 29 RNA:RNA interactions were detected in *E. coli* by Liu *et al.*, mainly involving rRNAs and tRNAs, reflecting the absence of rRNA removal before library construction. Consequently, the deepness of sequencing was probably too low to detect any sRNA-mRNA interactions.

In view of previously cited approaches, we are far from visualizing the whole cellular interactome in bacteria. However, the comparison of recently developed technologies, notably applied to the study of eukaryotic or viral RNA:RNA interactions will certainly help to achieve this goal. Indeed LIGR-Seq, PARIS, SPLASH and COMRADES rely on the same strategy of ligation between chemically cross-linked RNAs with psoralen derivates (such as AMT). LIGR-Seq method is based on the removal of non-ligated RNAs by enzymatic digestions (such as modified CLASH), while other methods include an additional purification step either by gel enrichment or pull-down of cross-linked RNA hybrids (for review see Schönberger *et al.* [91]). The main advantage of this purification step is to enrich true cross-linked RNA:RNA

interactions and to decrease the level of false positives, resulting in an increase of specificity. After proximity ligation, crosslinks are reversed at 254 nm, and RNA samples are subjected to RNA sequencing. However, due to the weak yield of crosslink and poor ligation efficiency, not all RNA:RNA interactions are revealed. Moreover, AMT has a preference to intercalate into opposite uracil bases. Thus, binding sites with a high proportion of GC residues can be underrepresented. Interestingly, in LIGR-seq method, the circRNA Ligase is favored over the T4 RNA Ligase to allow reaction at high temperature (60°C instead of 16°C), which reduces unspecific RNA:RNA interaction and ligation *in vitro*. Yet it will promote circularization of RNA, which might be excluded in the downstream steps of the library preparation. A good compromise would be to use another RNA Ligase operating at high temperature. The method used to prepare cDNA libraries is also a critical parameter. The adapter ligation method is more suitable than random priming to facilitate the mapping of both chimera extremities and identify interacting partners. In the same vein, samples fragmentation should be avoided. The development of a technology inspired by previously mentioned methods and taking into account suggested improvements will definitively help defining the whole RNA:RNA interactome.

3.4. Validation of candidates

Global analysis of RNA interactome raise many potential RNA:RNA pairs that need to be validated both *in vitro* and *in vivo* in order to decipher their functions. Jagodnik *et al.* (2017) reviewed extensively all the techniques used in the field [92]. Especially, the binding sites of recognition of sRNAs with their mRNA targets can be analyzed by gel retardation assays (also named electrophoretic mobility shift assay, EMSA). Chemicals or enzymatic probing can be used to examine the structure of regulatory complexes. Finally, the mechanisms of regulation can be characterized at the molecular level by analyzing the impact of the RNA:RNA complexes on RNA stability (Northern blot, rifampicin assays) and translation (toe-print, Western blot, gene reporter fusions). Here, we only focused on two recent technological advances that enable to measure at the genomic scale either RNA half-life or translation efficiency.

The analysis of variations in RNA decay rate in absence/presence of the studied sRNA not only helps understanding regulatory mechanisms, but also scouting new targets. To evaluate the dynamics of RNA transcription, processing and decay at the transcriptome-wide level, metabolic labeling experiments coupled with RNA sequencing have proven to be the *de facto* standard approaches [93]. In most protocols, 4-thiouridine (4sU) label, which is incorporated into newly synthesized transcripts, but not into pre-existing molecules, helps distinguishing both RNA populations. Remarkably, pulse-chase experiments allow the

determination of both the speed of transcription and RNA half-life. For this purpose, two main approaches are used: (1) an enrichment of the metabolically labeled RNAs by biochemical separation is achieved before library preparation. After total RNA isolation, thiol groups of 4sU-labelled RNAs are biotinylated and, thus efficiently purified. This approach, named 4sU-seq, has become quite popular in RNA biology, as exemplified by the increasing number of studies (see [94] for review). (2) 4sU nucleotides are converted into cytosines, which can be observed as point mutations in RNA-seq data (T to C transitions) allowing the partition during RNA-seq analysis of the newly synthetized transcripts from the non-mutated ones [95]. Current protocols are similar to the ones developed for bisulfite sequencing, which also uses chemically induced mutations to provide insight into RNA or DNA methylation. Here, the absence of any biochemical separation makes this method more accessible due to lower input amounts and a less laborious protocol. Although not yet used to study the impact of sRNAs on global mRNA turnover, these promising approaches, analyzing the kinetics of transcription and RNA decay, could lead to a more comprehensive understanding of the sRNA-dependent regulation.

Ribosome profiling experiments (Ribo-seq; [96]) precisely and quantitatively measure the translation rate of mRNAs in the cell, providing the opportunity to address the effect of sRNAs on the translation of mRNA targets. Commonly, sRNAs impact translation at the initiation steps either by repressing ribosome loading (occlusion of the Shine-Dalgarno sequence and/or start codon) or by stimulating it. In the latter case, the sRNA binding to an upstream region of the targeted transcript promotes secondary structure rearrangements around the ribosome binding site, which becomes more accessible. Ribo-seq method includes the purification of ribosomes bound to actively translated mRNAs from cell lysates. After RNase treatment, the ~30 nts nuclease-protected mRNA fragments are sequenced, enabling the determination of ribosome positions on transcripts. Wang and colleagues [97] have adopted Ribo-seq method to experimentally identify regulatory targets of RyhB sRNA in *E. coli*. In addition to validate previously characterized targets, Ribo-seq increased the regulatory targetome of RyhB by unveiling novel mRNA partners. Remarkably, Ribo-seq coupled to transcriptomic analysis gives a rather complete overview of all RNAs regulated by an sRNA at the post-transcriptional and/or translational levels. Still, observed results are not systematically a consequence of direct interactions. Another valuable advantage of this approach is to identify bifunctional sRNAs encoding small open reading frame (ORF) as published recently [98].

CONCLUSION

The recent and considerable evolution in high-throughput sequencing technologies has not only uncovered knowledge on sRNAs characterization and expression, but also on their cellular functions. This enables to build vast and complex regulatory networks integrating

perceived stimuli and the impact on bacterial physiology. To go further, cell-to-cell diversity should be considered. Indeed, heterogeneous behavior between bacterial individuals is usually observed, notably in specific conditions such as biofilm or dormancy, providing a benefit to adapt more quickly to environmental changes. Until now, a unique approach addresses it by combining super-resolution imaging with fluorescence *in situ* hybridization (smFISH). This notably revealed the subcellular localization of a sRNA:mRNA complex and the pairing kinetic in fixed cells [99]. Moreover, very recently Kannaiah and colleagues (2019) [100] were able to perform a transcriptomic analysis of different cellular compartments (cytosolic, membrane and poles). They noticed that several Hfq-dependent sRNAs were specifically enriched at the poles together with their mRNA targets. The study revealed that not only the direct interaction between the sRNA and its targets and their relative stoichiometry should be considered but also their cellular co-localization to define regulatory functions.

In the future, continuous progresses in the sensitivity of RNAseq technologies will definitely allow analysis of the expression profiling of sRNAs at single-cell level and unravel their targetome. This will certainly explain cell-to-cell variation within a cell population and help revealing a target hierarchy.

ACKNOWLEDGMENTS

We thank Pascale Romby (PR) for helpful advice and discussions. We also thank the Integrative Molecular and Cellular Biology (IMCBio) graduate school.

FUNDING

This work was supported by the “Agence Nationale de la Recherche” (ANR, Grant ANR-16-CE11-0007-01, RIBOSTAPH, and ANR-18-CE12-0025-04, CoNoCo, to PR). It has also been published under the framework of the labEx NetRNA ANR-10-LABX-0036 and of ANR-17-EURE-0023 (to PR), as funding from the state managed by ANR as part of the investments for the future program. DL was supported by the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 753137-SaRNAReg. ED was supported by the “Fondation pour la Recherche Médicale” (FDT201904007957).

REFERENCES

[1] M.C. Carrier, D. Lalaouna, E. Masse, Broadening the Definition of Bacterial Small RNAs: Characteristics and Mechanisms of Action, Annual review of microbiology, 72 (2018) 141-161.
[2] A. Brosse, M. Guillier, Bacterial Small RNAs in Mixed Regulatory Networks, Microbiology spectrum, 6 (2018).
[3] Y. Chao, K. Papenfort, R. Reinhardt, C.M. Sharma, J. Vogel, An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs, Embo j, 31 (2012) 4005-4019.

[4] D. Lalaouna, M.C. Carrier, S. Semsey, J.S. Brouard, J. Wang, J.T. Wade, E. Masse, A 3' external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise, Mol Cell, 58 (2015) 393-405.
[5] P. Mandin, S. Gottesman, Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA, Embo j, 29 (2010) 3094-3107.
[6] K. Koeppen, T.H. Hampton, M. Jarek, M. Scharfe, S.A. Gerber, D.W. Mielcarz, E.G. Demers, E.L. Dolben, J.H. Hammond, D.A. Hogan, B.A. Stanton, A Novel Mechanism of Host-Pathogen Interaction through sRNA in Bacterial Outer Membrane Vesicles, PLoS Pathog, 12 (2016) e1005672.
[7] D. Lalaouna, J. Baude, Z. Wu, A. Tomasini, J. Chicher, S. Marzi, F. Vandenesch, P. Romby, I. Caldelari, K. Moreau, RsaC sRNA modulates the oxidative stress response of Staphylococcus aureus during manganese starvation, Nucleic Acids Research, (2019).
[8] A. Smirnov, C. Schneider, J. Hor, J. Vogel, Discovery of new RNA classes and global RNA-binding proteins, Curr Opin Microbiol, 39 (2017) 152-160.
[9] R.F. Dos Santos, C.M. Arraiano, J.M. Andrade, New molecular interactions broaden the functions of the RNA chaperone Hfq, Curr Genet, (2019).
[10] M. Raina, A. King, C. Bianco, C.K. Vanderpool, Dual-Function RNAs, Microbiology spectrum, 6 (2018) 10.1128/microbiolspec.RWR-0032-2018.
[11] M. Miyakoshi, Y. Chao, J. Vogel, Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA, EMBO J, 34 (2015) 1478-1492.
[12] T. Romeo, C.A. Vakulskas, P. Babitzke, Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems, Environmental microbiology, 15 (2013) 313-324.
[13] K. Bharwad, S. Rajkumar, Rewiring the functional complexity between Crc, Hfq and sRNAs to regulate carbon catabolite repression in Pseudomonas, World journal of microbiology & biotechnology, 35 (2019) 140.
[14] J. Vogel, C.M. Sharma, How to find small non-coding RNAs in bacteria, Biological chemistry, 386 (2005) 1219-1238.
[15] J. Sridhar, P. Gunasekaran, Computational small RNA prediction in bacteria, Bioinformatics and biology insights, 7 (2013) 83-95.
[16] R.K. Carroll, A. Weiss, W.H. Broach, R.E. Wiemels, A.B. Mogen, K.C. Rice, L.N. Shaw, Genome-wide Annotation, Identification, and Global Transcriptomic Analysis of Regulatory or Small RNA Gene Expression in Staphylococcus aureus, MBio, 7 (2016) e01990-01915.
[17] S.R. Pernitzsch, C.M. Sharma, Transcriptome complexity and riboregulation in the human pathogen Helicobacter pylori, Front Cell Infect Microbiol, 2 (2012) 14.
[18] M. Wang, J. Fleming, Z. Li, C. Li, H. Zhang, Y. Xue, M. Chen, Z. Zhang, X.E. Zhang, L. Bi, An automated approach for global identification of sRNA-encoding regions in RNA-Seq data from Mycobacterium tuberculosis, Acta Biochim Biophys Sin (Shanghai), 48 (2016) 544-553.
[19] D. Sinha, K. Zimmer, T.A. Cameron, D.B. Rusch, M.E. Winkler, N.R. De Lay, Redefining the Small Regulatory RNA Transcriptome in Streptococcus pneumoniae Serotype 2 Strain D39, J Bacteriol, 201 (2019).
[20] O.A. Soutourina, M. Monot, P. Boudry, L. Saujet, C. Pichon, O. Sismeiro, E. Semenova, K. Severinov, C. Le Bouguenec, J.Y. Coppee, B. Dupuy, I. Martin-Verstraete, Genome-wide identification of regulatory RNAs in the human pathogen Clostridium difficile, PLoS Genet, 9 (2013) e1003493.
[21] M. Gomez-Lozano, R.L. Marvig, S. Molin, K.S. Long, Genome-wide identification of novel small RNAs in Pseudomonas aeruginosa, Environmental microbiology, 14 (2012) 2006-2016.
[22] M.P. Grull, L. Pena-Castillo, M.E. Mulligan, A.S. Lang, Genome-wide identification and characterization of small RNAs in Rhodobacter capsulatus and identification of small RNAs affected by loss of the response regulator CtrA, RNA Biol, 14 (2017) 914-925.
[23] G. Pei, T. Sun, S. Chen, L. Chen, W. Zhang, Systematic and functional identification of small non-coding RNAs associated with exogenous biofuel stress in cyanobacterium Synechocystis sp. PCC 6803, Biotechnology for biofuels, 10 (2017) 57.

[24] C. del Val, E. Rivas, O. Torres-Quesada, N. Toro, J.I. Jimenez-Zurdo, Identification of differentially expressed small non-coding RNAs in the legume endosymbiont Sinorhizobium meliloti by comparative genomics, Mol Microbiol, 66 (2007) 1080-1091.
[25] R. Frantz, L. Teubner, T. Schultze, L. La Pietra, C. Muller, K. Gwozdzinski, H. Pillich, T. Hain, M. Weber-Gerlach, G.D. Panagiotidis, A. Mostafa, F. Weber, M. Rohde, S. Pleschka, T. Chakraborty, M. Abu Mraheil, The secRNome of Listeria monocytogenes Harbors Small Noncoding RNAs That Are Potent Inducers of Beta Interferon, MBio, 10 (2019).
[26] B. Tjaden, De novo assembly of bacterial transcriptomes from RNA-seq data, Genome biology, 16 (2015) 1.
[27] S.H. Yu, J. Vogel, K.U. Forstner, ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes, GigaScience, 7 (2018).
[28] L. Pena-Castillo, M. Gruell, M.E. Mulligan, A.S. Lang, Detection of bacterial small transcripts from RNA-Seq data: A comparative assessment, Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 21 (2016) 456-467.
[29] S. Leonard, S. Meyer, S. Lacour, W. Nasser, F. Hommais, S. Reverchon, APERO: a genome-wide approach for identifying bacterial small RNAs from RNA-Seq data, Nucleic Acids Res, (2019).
[30] D. Tillett, B.P. Burns, B.A. Neilan, Optimized rapid amplification of cDNA ends (RACE) for mapping bacterial mRNA transcripts, BioTechniques, 28 (2000) 448, 450, 452-443, 456.
[31] M. Fromont-Racine, E. Bertrand, R. Pictet, T. Grange, A highly sensitive method for mapping the 5' termini of mRNAs, Nucleic Acids Res, 21 (1993) 1683-1684.
[32] M.F. Carey, C.L. Peterson, S.T. Smale, The Primer Extension Assay, Cold Spring Harbor Protocols, 2013 (2013) pdb.prot071902.
[33] R.A. Fekete, M.J. Miller, D.K. Chattoraj, Fluorescently labeled oligonucleotide extension: a rapid and quantitative protocol for primer extension, BioTechniques, 35 (2003) 90-94, 97-98.
[34] C.M. Sharma, J. Vogel, Differential RNA-seq: the approach behind and the biological insight gained, Curr Opin Microbiol, 19 (2014) 97-105.
[35] L. Ettwiller, J. Buswell, E. Yigit, I. Schildkraut, A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome, BMC Genomics, 17 (2016) 199.
[36] N. Innocenti, M. Golumbeanu, A. Fouquier d'Herouel, C. Lacoux, R.A. Bonnin, S.P. Kennedy, F. Wessner, P. Serror, P. Bouloc, F. Repoila, E. Aurell, Whole-genome mapping of 5' RNA ends in bacteria by tagged sequencing: a comprehensive view in Enterococcus faecalis, Rna, 21 (2015) 1018-1030.
[37] C.M. Sharma, S. Hoffmann, F. Darfeuille, J. Reignier, S. Findeiss, A. Sittka, S. Chabas, K. Reiche, J. Hackermuller, R. Reinhardt, P.F. Stadler, J. Vogel, The primary transcriptome of the major human pathogen Helicobacter pylori, Nature, 464 (2010) 250-255.
[38] N. Heidrich, S. Bauriedl, L. Barquist, L. Li, C. Schoen, J. Vogel, The primary transcriptome of Neisseria meningitidis and its interaction with the RNA chaperone Hfq, Nucleic Acids Res, 45 (2017) 6147-6167.
[39] A.N. Luck, B.E. Slatko, J.M. Foster, Removing the needle from the haystack: Enrichment of Wolbachia endosymbiont transcripts from host nematode RNA by Cappable-seq, PLoS One, 12 (2017) e0173186.
[40] J.G. Bird, Y. Zhang, Y. Tian, N. Panova, I. Barvik, L. Greene, M. Liu, B. Buckley, L. Krasny, J.K. Lee, C.D. Kaplan, R.H. Ebright, B.E. Nickels, The mechanism of RNA 5' capping with NAD+, NADH and desphospho-CoA, Nature, 535 (2016) 444-447.
[41] M.L. Winz, H. Cahova, G. Nubel, J. Frindert, K. Hofer, A. Jaschke, Capture and sequencing of NAD-capped RNA sequences with NAD captureSeq, Nature protocols, 12 (2017) 122-149.
[42] I.O. Vvedenskaya, J.G. Bird, Y. Zhang, Y. Zhang, X. Jiao, I. Barvik, L. Krasny, M. Kiledjian, D.M. Taylor, R.H. Ebright, B.E. Nickels, CapZyme-Seq Comprehensively Defines Promoter-Sequence Determinants for RNA 5' Capping with NAD<sup/>, Mol Cell, 70 (2018) 553-564.e559.

[43] J. Chen, T. Morita, S. Gottesman, Regulation of Transcription Termination of Small RNAs and by Small RNAs: Molecular Mechanisms and Biological Functions, Front Cell Infect Microbiol, 9 (2019) 201.
[44] H. Otaka, H. Ishikawa, T. Morita, H. Aiba, PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action, Proc Natl Acad Sci U S A, 108 (2011) 13059-13064.
[45] P. Mitra, G. Ghosh, M. Hafeezunnisa, R. Sen, Rho Protein: Roles and Mechanisms, Annual review of microbiology, 71 (2017) 687-709.
[46] E. Lioliou, C.M. Sharma, I. Caldelari, A.C. Helfer, P. Fechter, F. Vandenesch, J. Vogel, P. Romby, Global regulatory functions of the Staphylococcus aureus endoribonuclease III in gene expression, PLoS Genet, 8 (2012) e1002782.
[47] C.L. Kingsford, K. Ayanbule, S.L. Salzberg, Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake, Genome biology, 8 (2007) R22.
[48] P.P. Gardner, L. Barquist, A. Bateman, E.P. Nawrocki, Z. Weinberg, RNIE: genome-wide prediction of bacterial intrinsic terminators, Nucleic Acids Res, 39 (2011) 5845-5852.
[49] C.Q. Feng, Z.Y. Zhang, X.J. Zhu, Y. Lin, W. Chen, H. Tang, H. Lin, iTerm-PseKNC: a sequence-based tool for predicting bacterial transcriptional terminators, Bioinformatics, 35 (2019) 1469-1477.
[50] M. Di Salvo, S. Puccio, C. Peano, S. Lacour, P. Alifano, RhoTermPredict: an algorithm for predicting Rho-dependent transcription terminators based on Escherichia coli, Bacillus subtilis and Salmonella enterica databases, BMC Bioinformatics, 20 (2019) 117.
[51] D. Dar, M. Shamir, J.R. Mellin, M. Koutero, N. Stern-Ginossar, P. Cossart, R. Sorek, Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria, Science, 352 (2016) aad9822.
[52] R.A. Britton, T. Wen, L. Schaefer, O. Pellegrini, W.C. Uicker, N. Mathy, C. Tobin, R. Daou, J. Szyk, C. Condon, Maturation of the 5' end of Bacillus subtilis 16S rRNA by the essential ribonuclease YkqC/RNase J1, Mol Microbiol, 63 (2007) 127-138.
[53] X. Ju, D. Li, S. Liu, Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria, Nature microbiology, 4 (2019) 1907-1918.
[54] N. Peterman, A. Lavi-Itzkovitz, E. Levine, Large-scale mapping of sequence-function relations in small regulatory RNAs reveals plasticity and modularity, Nucleic Acids Res, 42 (2014) 12177-12188.
[55] N.R. Markham, M. Zuker, UNAFold: software for nucleic acid folding and hybridization, Methods Mol Biol, 453 (2008) 3-31.
[56] R. Lorenz, S.H. Bernhart, C. Honer Zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA Package 2.0, Algorithms for molecular biology : AMB, 6 (2011) 26.
[57] I.L. Hofacker, M. Fekete, P.F. Stadler, Secondary structure prediction for aligned RNA sequences, J Mol Biol, 319 (2002) 1059-1066.
[58] C.E. Hajdin, S. Bellaousov, W. Huggins, C.W. Leonard, D.H. Mathews, K.M. Weeks, Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots, Proc Natl Acad Sci U S A, 110 (2013) 5498-5503.
[59] P.R. Wright, M. Mann, R. Backofen, Structure and Interaction Prediction in Prokaryotic RNA Biology, Microbiology spectrum, 6 (2018).
[60] E. Mailler, J.C. Paillart, R. Marquet, R.P. Smyth, V. Vivet-Boudou, The evolution of RNA structural probing methods: From gels to next-generation sequencing, Wiley interdisciplinary reviews. RNA, 10 (2019) e1518.
[61] F. Righetti, A.M. Nuss, C. Twittenhoff, S. Beele, K. Urban, S. Will, S.H. Bernhart, P.F. Stadler, P. Dersch, F. Narberhaus, Temperature-responsive in vitro RNA structurome of Yersinia pseudotuberculosis, Proc Natl Acad Sci U S A, 113 (2016) 7237-7242.
[62] Z. Ignatova, F. Narberhaus, Systematic probing of the bacterial RNA structurome to reveal new functions, Curr Opin Microbiol, 36 (2017) 14-19.
[63] K.E. Watters, T.R. Abbott, J.B. Lucks, Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq, Nucleic Acids Res, 44 (2016) e12.

[64] D. Lalaouna, A. Eyraud, A. Devinck, K. Prevost, E. Masse, GcvB small RNA uses two distinct seed regions to regulate an extensive targetome, Mol Microbiol, 111 (2019) 473-486.
[65] M.B. Kery, M. Feldman, J. Livny, B. Tjaden, TargetRNA2: identifying targets of small regulatory RNAs in bacteria, Nucleic Acids Res, 42 (2014) W124-129.
[66] X. Ying, Y. Cao, J. Wu, Q. Liu, L. Cha, W. Li, sTarPicker: a method for efficient prediction of bacterial sRNA targets based on a two-step model for hybridization, PLoS One, 6 (2011) e22705.
[67] M. Mann, P.R. Wright, R. Backofen, IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions, Nucleic Acids Res, 45 (2017) W435-w439.
[68] P.R. Wright, J. Georg, M. Mann, D.A. Sorescu, A.S. Richter, S. Lott, R. Kleinkauf, W.R. Hess, R. Backofen, CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains, Nucleic Acids Res, 42 (2014) W119-123.
[69] A. Pain, A. Ott, H. Amine, T. Rochat, P. Bouloc, D. Gautheret, An assessment of bacterial small RNA target prediction programs, RNA Biol, 12 (2015) 509-513.
[70] D. Lai, I.M. Meyer, A comprehensive comparison of general RNA-RNA interaction prediction methods, Nucleic Acids Res, 44 (2016) e61.
[71] A.M. King, C.K. Vanderpool, P.H. Degnan, sRNA Target Prediction Organizing Tool (SPOT) Integrates Computational and Experimental Data To Facilitate Functional Characterization of Bacterial Small RNAs, mSphere, 4 (2019).
[72] D. Bronesky, E. Desgranges, A. Corvaglia, P. Francois, C.J. Caballero, L. Prado, A. Toledo-Arana, I. Lasa, K. Moreau, F. Vandenesch, S. Marzi, P. Romby, I. Caldelari, A multifaceted small RNA modulates gene expression upon glucose limitation in Staphylococcus aureus, Embo j, 38 (2019).
[73] A.E. Saliba, S. C Santos, J. Vogel, New RNA-seq approaches for the study of bacterial pathogens, Curr Opin Microbiol, 35 (2017) 78-87.
[74] D. Lalaouna, K. Prevost, G. Laliberte, V. Houe, E. Masse, Contrasting silencing mechanisms of the same target mRNA by two regulatory RNAs in Escherichia coli, Nucleic Acids Res, 46 (2018) 2600-2612.
[75] D. Lalaouna, A. Morissette, M.C. Carrier, E. Masse, DsrA regulatory RNA represses both hns and rbsD mRNAs through distinct mechanisms in Escherichia coli, Mol Microbiol, 98 (2015) 357-369.
[76] I.J. Silva, S. Barahona, A. Eyraud, D. Lalaouna, N. Figueroa-Bossi, E. Masse, C.M. Arraiano, SraL sRNA interaction regulates the terminator by preventing premature transcription termination of rho mRNA, Proc Natl Acad Sci U S A, 116 (2019) 3042-3051.
[77] D. Bronesky, E. Desgranges, A. Corvaglia, P. François, C.J. Caballero, L. Prado, A. Toledo-Arana, I. Lasa, K. Moreau, F. Vandenesch, S. Marzi, P. Romby, I. Caldelari, A dual sRNA in Staphylococcus aureus induces a metabolic switch responding to glucose consumption, bioRxiv, (2018).
[78] A. Tomasini, K. Moreau, J. Chicher, T. Geissmann, F. Vandenesch, P. Romby, S. Marzi, I. Caldelari, The RNA targetome of Staphylococcus aureus non-coding RNA RsaA: impact on cell surface properties and defense mechanisms, Nucleic Acids Res, 45 (2017) 6746-6760.
[79] D. Lalaouna, E. Desgranges, I. Caldelari, S. Marzi, Chapter Sixteen - MS2-Affinity Purification Coupled With RNA Sequencing Approach in the Human Pathogen Staphylococcus aureus, Methods in Enzymology, 612 (2018) 393-411.
[80] D. Lalaouna, K. Prevost, A. Eyraud, E. Masse, Identification of unknown RNA partners using MAPS, Methods, 117 (2017) 28-34.
[81] T. Rochat, C. Bohn, C. Morvan, T.N. Le Lam, F. Razvi, A. Pain, C. Toffano-Nioche, P. Ponien, A. Jacq, E. Jacquet, P.D. Fey, D. Gautheret, P. Bouloc, The conserved regulatory RNA RsaE down-regulates the arginine degradation pathway in Staphylococcus aureus, Nucleic Acids Res, 46 (2018) 8803-8816.
[82] S. Melamed, A. Peer, R. Faigenbaum-Romm, Y.E. Gatt, N. Reiss, A. Bar, Y. Altuvia, L. Argaman, H. Margalit, Global Mapping of Small RNA-Target Interactions in Bacteria, Mol Cell, 63 (2016) 884-897.

[83] S. Melamed, R. Faigenbaum-Romm, A. Peer, N. Reiss, O. Shechter, A. Bar, Y. Altuvia, L. Argaman, H. Margalit, Mapping the small RNA interactome in bacteria using RIL-seq, Nature protocols, 13 (2018) 1-33.
[84] S. Melamed, P.P. Adams, A. Zhang, H. Zhang, G. Storz, RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles, Mol Cell, (2019).
[85] S.A. Waters, S.P. McAteer, G. Kudla, I. Pang, N.P. Deshpande, T.G. Amos, K.W. Leong, M.R. Wilkins, R. Strugnell, D.L. Gally, D. Tollervey, J.J. Tree, Small RNA interactome of pathogenic E. coli revealed through crosslinking of RNase E, Embo j, 36 (2017) 374-387.
[86] J. Hor, J. Vogel, Global snapshots of bacterial RNA networks, Embo j, 36 (2017) 245-247.
[87] T. Morita, H. Aiba, RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria, Genes Dev, 25 (2011) 294-298.
[88] Y.F. Zhang, K. Han, C.E. Chandler, B. Tjaden, R.K. Ernst, S. Lory, Probing the sRNA regulatory landscape of P. aeruginosa: post-transcriptional control of determinants of pathogenicity and antibiotic susceptibility, Mol Microbiol, 106 (2017) 919-937.
[89] K. Han, B. Tjaden, S. Lory, GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation, Nature microbiology, 2 (2016) 16239.
[90] T. Liu, K. Zhang, S. Xu, Z. Wang, H. Fu, B. Tian, X. Zheng, W. Li, Detecting RNA-RNA interactions in E. coli using a modified CLASH method, BMC Genomics, 18 (2017) 343.
[91] B. Schonberger, C. Schaal, R. Schafer, B. Voss, RNA interactomics: recent advances and remaining challenges, F1000Research, 7 (2018).
[92] J. Jagodnik, A. Brosse, T.N. Le Lam, C. Chiaruttini, M. Guillier, Mechanistic study of base-pairing small regulatory RNAs in bacteria, Methods, 117 (2017) 67-76.
[93] A. Uvarovskii, I.S. Naarmann-de Vries, C. Dieterich, On the optimal design of metabolic RNA labeling experiments, PLoS computational biology, 15 (2019) e1007252.
[94] L. Wachutka, J. Gagneur, Measures of RNA metabolism rates: Toward a definition at the level of single bonds, Transcription, 8 (2017) 75-80.
[95] T. Yamada, N. Akimitsu, Contributions of regulated transcription and mRNA decay to the dynamics of gene expression, Wiley interdisciplinary reviews. RNA, 10 (2019) e1508.
[96] N.T. Ingolia, S. Ghaemmaghami, J.R. Newman, J.S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling, Science, 324 (2009) 218-223.
[97] J. Wang, W. Rennie, C. Liu, C.S. Carmack, K. Prevost, M.P. Caron, E. Masse, Y. Ding, J.T. Wade, Identification of bacterial sRNA regulatory targets using ribosome profiling, Nucleic Acids Res, 43 (2015) 10308-10320.
[98] J. Weaver, F. Mohammad, A.R. Buskirk, G. Storz, Identifying Small Proteins by Ribosome Profiling with Stalled Initiation Complexes, mBio, 10 (2019).
[99] J. Fei, D. Singh, Q. Zhang, S. Park, D. Balasubramanian, I. Golding, C.K. Vanderpool, T. Ha, RNA biochemistry. Determination of in vivo target search kinetics of regulatory noncoding RNA, Science, 347 (2015) 1371-1374.
[100] S. Kannaiah, J. Livny, O. Amster-Choder, Spatiotemporal Organization of the E. coli Transcriptome: Translation Independence and Engagement in Regulation, Mol Cell, 76 (2019) 574-589.e577.

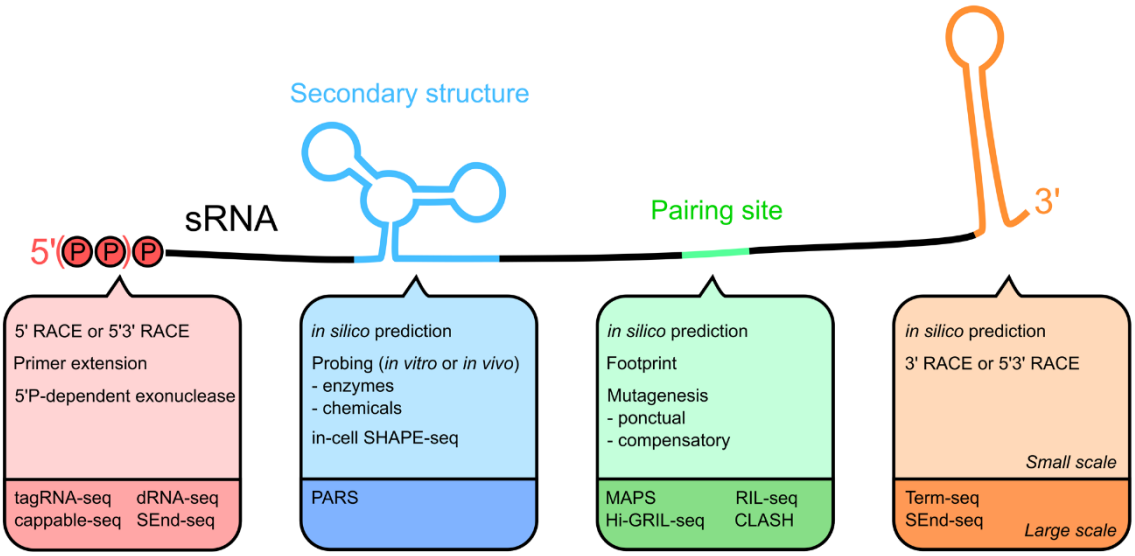

Figure 1. Schematic representation of the main methods (either on a single RNA, upper panels, or globally, lower panels) used to characterize an sRNA.

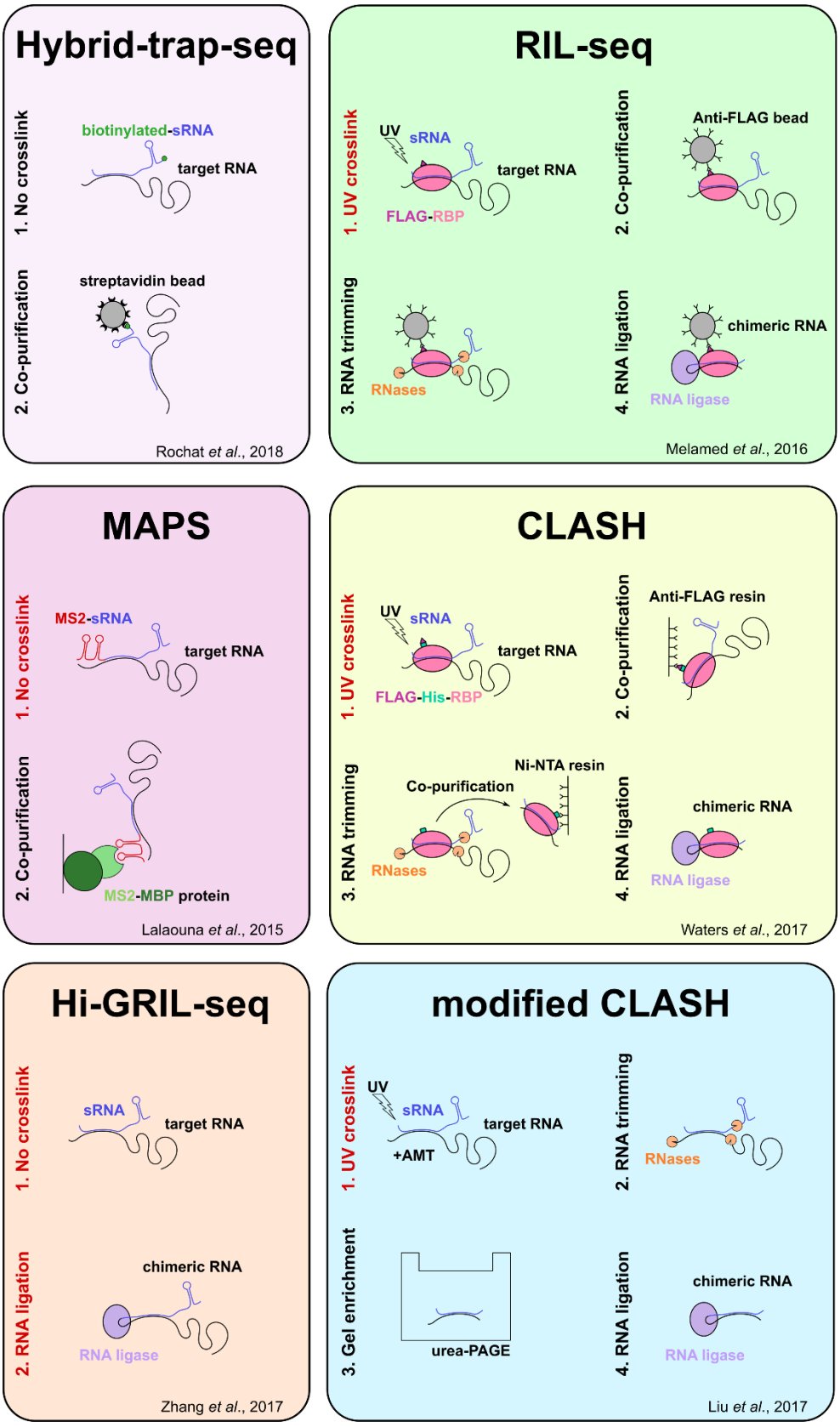

Figure 2. Comparison of Hybrid-trap-Seq, MAPS, RIL-seq, CLASH, Hi-GRIL-seq and modified CLASH approaches. Steps indicated in red are performed *in vivo*, while those in black are performed *in vitro*. More details are provided in section 3.