THE UNIVERSITY
of EDINBURGH

THE UNIVERSITY of EDINBURGH

The role of small regulatory RNA networks in controlling adaptive responses in *Escherichia coli*

Ira Iosub

Thesis submitted for the degree of
Doctor of Philosophy

The University of Edinbugh,
October, 2017

DECLARATION

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise.The work has not been submitted for any other degree or professional qualification.

Ira Iosub,

October, 2017

ACKNOWLEDGEMENTS

I am grateful to many people for the support I got throughout the years of my PhD.

Most of all, I would like to thank my supervisor, Sander Granneman. I could not have gotten here without your relentless support, encouragement, and inspiring passion for science that you are so eager to share! Thank you for being understanding, and for giving me the chance to grow and find a productive rhythm. Thank you for being so much more optimistic than I am, and that after every chat we had I felt more confident. Under your mentoring I have become a better scientist, and a better person, and for that I will be forever grateful.
I look forward to the next few exciting months of continuing the research and to the discussions we will have about it!

I would also like to thank the current and past members of the Granneman lab team. Stuart! I was and still am so happy you joined the team! Thank you for being a great, caring friend and amazing co-worker! Thanks for all the chats we had, scientific or not! And of course, thank you for helping out with experimental work.
Thank you Elena for being a true friend I've made since I started my PhD studies. Thank you for being there, for always listening, for the good advice, for sometimes knowing better than myself what is good for me. You are missed.
Thank you Ralph and Rob for the training you have given me in my first few months in the lab. Ralph, everything I have learnt from you has become something I feel secure on - thank you for being a great teacher. Rob, thank you for encouraging me to challenge 'accepted views'.
I am grateful to David Tollervey, Bill Earnshaw and Atlanta Cook for being my thesis committee and watching over me these years. Thank you for the valuable advice and for exposing me to diverse scientific angles I could view my research from. I also thank my collaborator, Gabriela Viero, for the polysome profiling experiments and Jai Tree for providing strains and constructs, and for the discussions about the data. I am grateful to the Wellcome Trust for funding my PhD. Would also like to thank the students I supervised, Julia Nieken and Lutfi Othman for the work they contributed and for having me as their supervisor.
I am grateful to my family for always helping in every way that they could. Thank you, Alejandro for entering my life. Maria, Oz and Aiste, thank you for your friendship, for meeting up and for the lovely chats.

LAY SUMMARY

Bacteria are unicellular organisms with exquisite ability to rapidly adapt to changes in the environment. This ability relies on their capacity to precisely tailor their response to different environmental stress factors. At a cellular level, this response represents the capacity of the bacteria to “know” which bits from their genetic information/information encoded in their genomes would be most effective in their stress coping strategy.

We used *E.coli,* a model bacterial organism, to further investigate how bacteria “know” what genetic information to use and how it regulates its use. The genetic information is represented by molecules called messenger RNA (mRNA). Imagine the mRNA as a string in a violin. As one would obtain the ideal note by rough and/or fine tuning the violin, so the bacterial cell regulates its mRNAs through rough and fine tuning factors. Although it is known that the fine-tuning is fundamental for cell survival, it is still largely unknown how it takes place. Key players are the small regulatory RNAs, which directly interact with mRNAs and decide their fate. For a violin string, not only the tuning, but also the timing is essential to ‘sound right’, as it is for effective responses in bacteria. Bacteria coordinate multiple responses simultaneously, and both the amplitude of tuning and the timing are tightly interconnected between these responses. We have shown here that these circuits are more interconnected, and richer in fine-tuning factors and ways of tuning than previously considered. I have shown in detail for a new small RNA, how it could act to help cells control a stress-response, while optimising the uptake of specific nutrients from the environment.

ABSTRACT

Microorganisms are exposed to constantly changing environments, and consequently have evolved mechanisms to rapidly adapt their physiology upon stress imposition. These adaptive responses are coordinated through the rewiring of gene expression via complex networks that control the transcriptional program and the activity of post-transcriptional regulators. Although transcription factors primarily determine which genes are expressed, post-transcriptional regulation has a major role in fine-tuning the dynamics of gene expression.

Post-transcriptional control is exerted by RNA-binding proteins and small regulatory RNAs (sRNAs) that bind to mRNA targets and modulate their synthesis, degradation and translation efficiency. In *Escherichia coli,* sRNAs associated with an RNA chaperone, Hfq, are key post-transcriptional regulators, yet the functions of most of these sRNAs are still unknown. The first step in understanding the roles of sRNAs in regulating gene expression is to identify their targets. To generate transcriptome-wide maps of Hfq-mediated sRNA-mRNA binding, we applied CLASH (cross-linking, ligation and sequencing of hybrids), a method that combines *in vivo* capture of RNA-RNA interactions, high-throughput sequencing and computational analyses, in *E. coli*. We uncovered thousands of dynamic growth-stage dependent association of Hfq to sRNAs and mRNAs. The latter confirmed known sRNA-target pairs and identified additional targets for known sRNAs, as well as novel sRNAs in various genomic features along with their targets. These data significantly expand our knowledge of the sRNA-target interaction networks in *E.coli.* In particular, the Hfq CLASH data indicated 3'-UTRs of mRNAs as major reservoirs of sRNAs, and the utilization of these may be more common than anticipated. Our findings also provide mechanistic insights that ensue from the identification of tens of sRNA-sRNA interactions that point to extensive sponging activity among regulatory RNAs: many sRNAs appear to be able to interact and repress the functions of other base-pairing sRNAs.

We validated and highlighted the biological significance of some of the CLASH results by characterizing a 3'-UTR derived sRNA, MdoR (*mal*-dependent OMP repressor). This sRNA emerges by processing of the last transcript of *malEFG* polycistron, encoding components of maltose transport system. We found MdoR directly downregulates several major porins, whilst derepressing the maltose-specific porin LamB via destabilization of its inhibitor, MicA, likely by a sponging mechanism. Physiologically, MdoR contributes to the remodelling of envelope composition and links nutrient sensing to envelope stress responses during maltose assimilation. MdoR is a clear example of how cells integrate circuitry through multiple networks as part of their adaptive responses and how the CLASH methodology can help expand our understanding of sRNA-based regulation.

Table of Contents

ABBREVIATIONS

5'-P: 5'-monophosphate

5'-PPP: 5'-triphosphate

A: Adenine

ATP: Adenosine Triphosphate

bp: base-pair

C-source: carbon source

cDNA: Complementary DNA

CsrA: Carbon storage regulator A

C: Cytosine

CLASH: Cross-linking, ligation and analysis of hybrids

CLIP: Cross-linking immunoprecipitation

CRAC: Cross-linking and analysis of cDNA

ESR: Extracytoplasmic stress response

FFL: Feed-forward loop

G: Guanine

HTF tag: His6-TEV-FLAG3 tag

IM: inner membrane

LB: Lysogeny broth

mRNA: messenger RNA

nt: nucleotide

OD_{600}: Optical Density measured at 600 nm wavelength

OM: outer membrane

OMP: outer membrane protein

RBP: RNA binding protein

RBS: ribosomal binding site

RIL-seq: RNA interaction by ligation and sequencing

RIP-seq: RNA Immunoprecipitation coupled to sequencing

RNase E: ribonuclease E

RNAP: RNA Polymerase

RT: reverse transcription

rRNA: ribosomal RNA

SD sequence: Shine Dalgarno sequence

SM: seed mutant

sRNA: small non-coding RNA

sfGFP: superfolder Green Fluorescent Protein

σ^{70}: housekeeping sigma factor

σ^{E}: sigma E alternative sigma factor

σ^{S}: general stress alternative sigma factor

TCS: two-component system

TF: Transcription factor

tRNA: transfer RNA

U: Uridine

UTR: untranslated region.

1. Introduction

1.1. Adaptive responses in microorganisms

Cellular stress constitutes a variety of processes triggered by sudden or sustained shifts from the usual cellular conditios and homeostasis. Cells respond to stress by counteracting the perturbation, repairing the damage and protecting the cell. Microorganisms are constantly exposed to a wide range of environmental changes. In their natural environments, organisms are often challenged with nutrient deprivation, temperature shifts, pH changes and other stress-inducing factors (Fig. 1.1). To survive, cells have evolved mechanisms to respond to these conditions by adapting their physiology. Among microorganisms, bacteria display remarkably fast adaptive responses and resilience to stress imposition, that not only provide a competitive advantage, but that also might define the difference between survival and death. In the case of pathogens, fast acclimation to the new host is a determining factor of success, and manifests itself by host invasion and persistence, survival in nutrient-poor conditions, increased expression of virulence factors, and evasion of host defence systems. Additionally, in response to antibiotics, bacteria trigger the activation of intrinsic antibiotic resistance mechanisms, including expression of efflux pumps. In a number of bacterial species, it has been established that virulence and antibiotic resistance mechanisms are intimately connected with stress responses (Flint et al., 2016; Geiger et al., 2012).

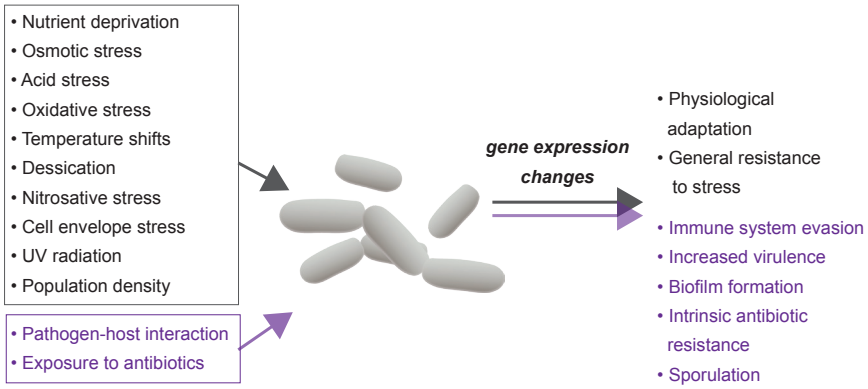

Fig. 1.1: Stress-inducing agents (left) and the main mechanism and physiological consequences of adaptive responses in bacteria; the text in purple and purple arrows indicate components often encountered in pathogens.

Stress-adaptive responses are coordinated by changes in gene expression. Bacteria employ complex strategies to attune gene expression with the physiological needs dictated by external and internal signals. Gene regulation is wired through gene regulatory networks that receive environmental signals and define the amplitude and timing of the responses, by integrating multiple inputs and outputs of the signalling (Sections 1.9-1.10). Stress-related gene expression programs are characterized by specific transcriptional and post-transcriptional mechanisms that need to be balanced with growth-related gene expression programs, to facilitate efficient adaptation to changing environments. Transcription primarily dictates which genes are expressed, but post-transcriptional regulation is key in shaping the dynamics of gene expression changes (Sections 1.2 and 1.10).

The post-transcriptional control component that will be the focus of this work is small RNA-mediated regulation (Section 1.3). Small non-coding RNAs are a diverse class of post-transcriptional regulators that affect many aspects of bacterial physiology, especially in relation to adaptation to stress. The central organism of this work is *Escherichia coli*, a Gram-negative gamma proteobacterium belonging to the Enterobacteriacea family and the most extensively studied bacterial model system. This makes *E. coli* an ideal organism for unravelling novel regulatory mechanisms or components, and facilitates the integration of new findings in the broader context of its physiology, to contribute to a global view of gene expression control. However, I will often refer and make parallels to studies performed in *Salmonella*, not only due to their similarity in the core genome and resemblance in post-transcriptional regulons, but also because many studies performed in this organism contributed immensely to the current knowledge of enterobacterial gene expression regulation. Noteworthy, the processes described in any bacterial species, regardless of their similarity in the core genomes, may have very different implications even for members of the same species, especially when the pan-genome is taken into account. The pan-genome includes the core genome (present in all representatives), accessory genome (present in several strains and even genes unique to single strains (Mira et al., 2010). This variable set of genes is acquired by lateral transfer between strains or even isolates from different

species (Mira et al., 2010). The pan-genome becomes even more important when studying the regulation of adaptive responses, because it is, in fact, an adaptive genome, that includes key genes for survival in defined environments, virulence and antibiotic resistance (Mira et al., 2010). These types of genes are under tight postr-transcriptional control, thus the rewiring of gene expression for different strains within the same species and the pathways involved can differ significantly.

Ultimately, the scope of the work presented in this thesis (Chapters 3-4) is to contribute to a better understanding of small RNA-mediated regulation and small RNA networks, by providing insights into small RNA-mediated regulation dynamics and maps of small-RNA-target interactions (Chapter 3) and detailed characterization of a case in which two small RNA pathways are linked to provide a robust response to the growth conditions (Chapter 4). This Chapter will introduce in detail the key concepts for understanding the work discussed in the following Chapters, but also provide a framework by briefly introducing related concepts.

1.2. Regulatory strategies to coordinate gene expression in response to changing environments

Transcription of genomic information into messenger RNA (mRNA) and translation of mRNAs into proteins are the main layers of gene expression (Figure 1.2 A). Transcription primarily determines which genes are synthesised, however, additional post-transcriptional control at multiple levels is needed for shaping gene expression profiles. Due to the fundamental need to adjust gene expression in response to both intracellular and extracellular cues, both processes are subject to tight regulation. This section will overview the central players and strategies in the regulation of both processes in bacteria, with an emphasis on the post-transcriptional control and some of the advantages it poses during adaptive responses.

The role of the transcriptional program in coordinating gene expression

The control of expression of stress-related genes is initiated at the transcription level

(Fig. 1.2 B), a highly-controlled process. In prokaryotes, most genes are organized in operons, sets of adjacent functionally related genes that are co-transcribed. This architecture allows coordinated transcriptional activation in response to cellular needs, with a substantial metabolic gain (Sneppen et al., 2010).Transcription is carried out by the RNA polymerase (RNAP) holoenzyme (Fig. 1.2 B). Transcription in bacteria requires sigma factor proteins (σ). The sigma factor is a subunit that associates with core RNAP to form the active holoenzyme (Ishihama, 2000; Paget., 2015) by providing the specificity for binding of RNAP to the promoter sequences of genes and allowing transcription initiation (Fig. 1.2 B). *E. coli* has seven sigma subunits,

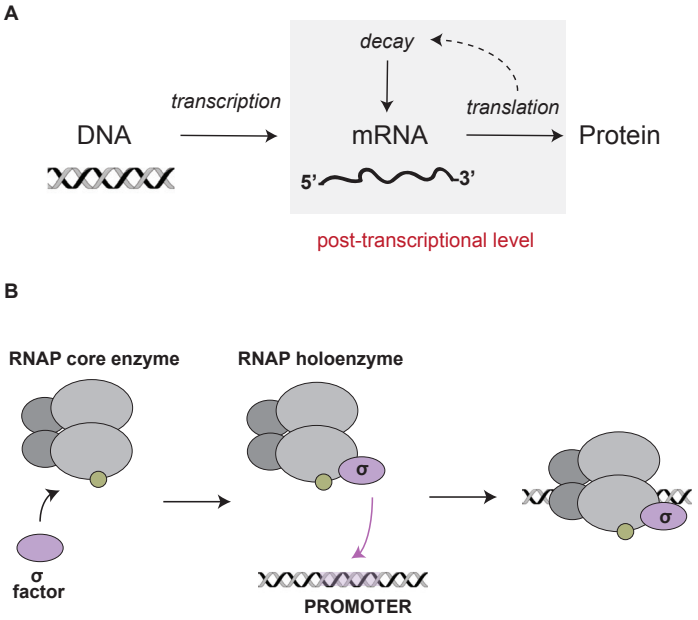

Fig. 1.3: (A) An abbreviated pathway of gene expression in bacteria. The grey shaded area indicates the post-transcriptional level of control; the dashed line indicates the link between the translation rate and decay (see text for details). (B) Transcription initiation requires sigma factors.The association of a sigma factor with the RNA Polymerase core enzyme provides the specificity of the RNAP binding to promoter sequences.

hence a total of seven functionally differentiated RNAP species, each recognizing a specific set of promoters (Ishihama, 2010; Maeda et al., 2000). Each sigma factor activates transcription of a defined set of stress-specific genes (e.g. σ^E for envelope stress; σ^{32} for heat-shock). The RNAP core is limiting and at relatively constant levels throughout growth, thus the identity of associated sigma factors is a major determinant of the pattern of gene expression, and sigma factors have been reported to compete

for binding to RNAP (Ishihama, 2010; Maeda et al., 2000). In conditions of stress, σ^{70}-RNAP (house-keeping holoenzyme) is outcompeted by RNAP associated with alternative sigma factors that transcribe stress-responsive genes.

The activity of RNAP is further modulated by transcription factors (TFs: ~300 in *E.coli*; Ishihama, 2014). TFs bind target DNA sequences, typically at sites near promoters, interact with promoter-bound RNAP and alter its activity and specificity. Thus, transcription factors can act as either activators or repressors.TFs contain one domain that functions as a signal sensor by ligand-binding or protein-protein interactions. Typically, the ligand is a physio-chemical signal or metabolite that carries the internal or environmental information (Martinez-Antonio et al., 2006; Balleza et al., 2009). The second domain of a TF is the responsive element, that directly interacts with DNA. Some TFs are part of two- or three-component systems (sensor and response regulators) that work as a unit. The topology of TFs allows rapid changes in transcription in response to specific environmental cues (Mascher et al., 2006; Ulrich et al., 2005).

Transcription elongation and termination are also highly-regulated steps. Transcription elongation can be regulated by RNAP pausing, that can cause attenuation or anti-termination (Washburn and Gottesman, 2015). In bacteria, transcription termination either requires the Rho factor, or is controlled by intrinsic terminator structures (Rho-independent terminators). Premature termination or its prevention are additional modes of gene expression regulation (Sedlyarova et al., 2016; Peters et al., 2013).

Additionally, chromosome organization and DNA-binding proteins that associate with DNA at non-specific sequences can control transcriptional activity epigenetically (Travers and Muskhelishvili, 2005).

Post-translational regulation: a major determinant of gene expression profiles

mRNA fate and protein output are further controlled by post-transcriptional

mechanisms, a mode of regulation heavily employed in stress-responses. In a wide range of organisms, global mRNA steady-states do not correlate well with protein levels, indicating a dominant role for post-transcriptional regulatory processes (Csardi et al., 2015; Schwanhausser et al., 2011; Maier et al., 2009; Picard, 2009). Post-transcriptional regulators include regulatory RNAs (riboregulators) or RNA binding proteins (RBPs) that modulate transcript stability and translation efficiency (Fig. 1.2). Most frequently, mRNA degradation and translation are regulated at their rate-limiting step, initiation. In many cases, cis- and trans-acting elements act in tandem to control gene expression (Van Assche et al., 2015). However, since in prokaryotes, transcription and translation are coupled, post-transcriptional regulators do in some cases influence RNA synthesis.

It has been increasingly recognized that post-transcriptional regulation of bacterial gene expression plays a key role in adaptive responses. Post-transcriptional control poses advantages in conditions that require rapid adaptation (detailed in Section 1.10) - compared to transcriptional regulation, this type of regulation does not require the relatively more resourceful synthesis of transcription factors. Moreover, many post-transcriptional regulators are short-lived or can be rapidly and transiently inactivated (e.g. by sponging; Section 1.6) or stabilized when needed, ensuring a controlled duration and amplitude of the response.

Additionally, whereas generally transcription factors regulate sets of functionally-related genes, post-transcriptional control can control the expression of a single gene or multiple unrelated genes, thereby allowing crosstalk and integration of different pathways (Section 1.9). However, it is the combined effect of transcriptional and post-transcriptional circuits that ultimately define cellular response curves.

Determinants of transcript stability

RNA processing and degradation have critical roles on the shaping of gene expression profiles. In order to understand how post-transcriptional regulators (Section 1.3) affect

mRNA degradation rates, we first have to inspect closer how transcript half-lives are modulated.

Compared to rRNAs and tRNAs, which are long-lived and functional for more generations, mRNA molecules are characterised by metabolic instability. Typical mRNA half-lives range from 1-2 minutes, although some mRNAs can have half-lives of up to 10-15 minutes (e.g. *ompA;* Hansen et al., 1994). Although mRNA half-life is uniquely and intrinsically determined by its primary sequence and secondary structure, controlled cellular processes such as translation rate (hence ribosome occupancy) and the binding of RBPs and other RNAs can significantly alter the fate of mRNAs (Radhakrisshnan and Green, 2016; Hui et al., 2014; Mackie, 2013; Joyce and Dreyfus, 1998; Wagner et al., 1994). The instability of mRNAs allows modulation of the amplitude and duration of the changes in gene expression in response to environmental changes to attain homeostasis.

In prokaryotes, controlled decay is monitored by endo- and exoribonucleases and the factors that recruit them (Hui et al.,2014; Bandyra et al.,2013). The key step in the major pathway of degradation of transcripts in *E. coli* is an initial cleavage by an endoribonuclease, usually RNase E (Carpousis et al., 2007; Section 1.5), followed by complete degradation by 3'-exoribonucleases and oligoribonucleases. This is aided by numerous other enzymes, such as RNA helicases and poly(A) polymerase (PAP) (Bandyra et al., 2013; Mackie et al., 2013; Lisitsky and Schuster., 1999). However, the RNA decay pathways are not universal as the same transcript can be turned over by different pathways depending on the growth conditions or during stress.

Consequences of regulating mRNA degradation on gene expression

Post-transcriptional regulation accounts for the fine-tuning of gene expression, and can greatly influence the kinetics of a response. This is clearly reflected by the effects of regulation on transcript stability. Controlled RNA decay (Hui et al., 2014; Mackie et al., 2013) has been shown to be determinant for the post-transcriptional

modulation of gene expression not only during growth, but also during adaptation to new environments (Arraiano et al., 2010). The steady-state abundance of a certain message is determined by the rates of transcription and degradation, therefore fine-tuning the balance between mRNA production and degradation can define transcriptome kinetics, in a condition-specific manner. For instance, models describing mRNA fate after the imposition of acute stress (Shalem et al., 2008) revealed that the concomitant increase in both transcription and degradation rates results in a more immediate increase in mRNA abundance (the half-maximal steady-state is reached

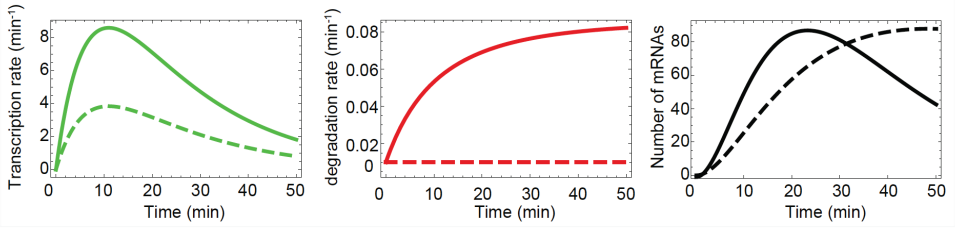

Fig. 1.4: The effects of activating both transcription and RNA decay on the kinetic profile of total RNA levels. The curves with a dashed line represent the dynamics of the response whereby only the transcription of a gene is induced, with no degradation. In both scenarios, the same steady state level of the RNA is achieved, but the activation of both transcription and decay allows that steady state to be reached sooner, resulting in faster responses. Figure courtesy of Nacho Molina.

faster), and is followed by a decrease in steady states. This ensures a much faster and transient response than by only switching transcription on and off (Fig. 1.3). This type of dynamics is advantageous in the regulation of stress-responsive genes, as it is not only rapid, but also reversible, hence it allows optimized activity of gene products during each phase of the response: shock, adaptation and recovery. Multiple studies reported the control of the stability of stress-responsive mRNAs (Molin et al., 2009; Garcia-Martinez et al., 2004), suggesting that most of these messages have a dynamic range of half-lives depending on the conditions (Dressaire et al., 2013; Selinger et al., 2003).

Regulation of translation as an additional layer of post-transcriptional control

Whereas the regulation of transcript levels is determinant in the number of mRNAs that can be potentially translated into proteins, it is the translation rate that ultimately

determines which proteins are expressed. Left unregulated, translation manifests as an amplification of transcriptional activity, i.e. a mRNA can be perpetually translated, and the output is dependent on translation rate and mRNA half-life. It is not surprising that this level of gene expression is tightly regulated at all steps (initiation, elongation, termination) and requires a large number of factors. The assembly of the protein synthesis machinery occurs during translation initiation, a process that brings together mRNA, initiator transfer RNA, and the ribosome into the 30S initiation complex (Simonetti et al., 2009; Kozak, 1999). The bacterial mRNA ribosome binding site (RBS) typically contains a polypurine stretch known as the Shine-Dalgarno (SD) sequence, located typically 4-5 nt upstream the start codon of mRNAs. SD sequence is required for the interaction between mRNA and the 30S ribosomal subunit, which is key for initiation of protein synthesis (Shine and Dalgarno, 1974; Chen et al., 1994). Translation initiation is the rate-limiting step in protein-synthesis, and is under tight post-transcriptional control (Duval et al., 2015). Since formation of the initiation complex is highly dependent on RNA-RNA interactions, the recruitment of the small ribosomal subunit can be modulated by mRNA secondary structure rearrangements, antisense RNA binding, or RBPs (Section 1.5).

1.3. Trans-acting small RNAs - a major class of riboregulators with prevalent roles in gene expression

The past few decades of research have established that the versatility of RNA expands beyond that as an information-carrying and structural molecule. This property ensues from its ability to form complex, yet flexible secondary structures and its ability to interact with other macromolecules (e.g RPBs), and mimic other nucleic acids. RNA-mediated regulation is a widespread phenomenon in all organisms. In bacteria, riboregulators are a structurally and functionally heterogeneous group of molecules that act in *cis*- or in *trans*- and that modulate gene expression or protein activity. These are typically non-coding RNAs, although some riboregulators may also encode for small proteins (Gimpel and Brantl, 2016; Wadler and Vanderpool, 2007).

This section will introduce the riboregulators that respond to environmental changes,

and not regulatory RNAs with house-keeping functions. The most extensive class of regulators, small non-coding RNAs (sRNAs) has the most prevalent effect in post-transcriptional networks, and will be the emphasis of this section.

RNA regulators that are part of the RNAs they regulate

In cis RNA regulators are structurally a part of the RNA they regulate, and are represented in bacteria by RNA thermosensors (Kortmann and Narberhaus, 2012) and riboswitches (Henkin, 2008), which sense and respond to temperature changes and nutrient availability, respectively. Their leader sequences fold into secondary structures (e.g. inhibitory hairpins that occlude the RBS in the case of RNA thermosensors) that undergo conformational changes upon temperature shifts or small-ligand binding, resulting in changes in expression of the downstream gene.

Small non-coding RNAs

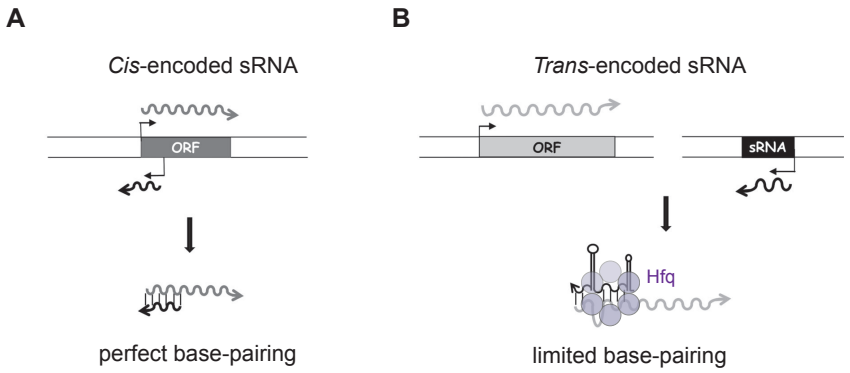

Fig. 1.5: Comparison between *cis*-encoded sRNAs and *trans*-encoded sRNAs. ORF indicates the open reading frame, arrows indicate transcriptional start sites (A) *Cis*-encoded sRNAs are encoded opposite their target mRNAs and have perfect complementarity with their target RNAs. (B) *Trans*-encoded sRNAs are transcribed from genomic loci separate from those encoding their target RNAs and have only partial complementarity, thus base-pairing is assisted by Hfq protein (purple). Figure adapted and modified from Prasse et al., 2013.

The largest class of riboregulators, the small RNAs (sRNAs) act by base-pairing with other RNAs, most frequently mRNAs. These riboregulators help cells adjust to environmental changes by modulating expression of key proteins. sRNAs can be

encoded either in cis, on the opposite DNA strand of the target, or trans, at a distinct locus in the genome relative to the targets (Fig. 1.3). Both types are produced as discrete RNAs that can base-pair with independently-transcribed targets.

Cis-encoded antisense RNAs

Cis-encoded sRNAs exhibit extended complementarity with their targets (Brantl, 2007) (Fig. 1.4 A), and interfere with their translation, induce premature transcription termination or promote their degradation (Georg and Hess, 2011; Thomason and Storz, 2010; Brantl, 2007). Cis-encoded antisense sRNAs are typically encoded in plasmids, transposons and phages, and more rarely within the bacterial chromosome (Brantl, 2007; Kawano et al., 2005). Two antisense sRNAs have roles in the repression of expression of toxic proteins (e.g. OhsC in *E. coli*; Fozo et al., 2008) or the uncoupling of expression of genes within the same operon (Tramonti et al., 2008). However, the perfect complementarity required for this type of regulation implies that the targets are unique, or in a limited number. Some *cis*-encoded sRNAs may act as dual regulators that also regulate mRNAs encoded in *trans*, similar to classical *trans*-encoded sRNAs (Thomason and Storz, 2010; Mandin and Gottesman, 2010; Jäger et al., 2012; Melamed et al., 2016).

Trans-encoded sRNAs

Trans-encoded sRNAs are the most abundant and diverse class of riboregulators, and together with the activity of TFs they shape gene expression profiles in response to specific extracellular and intracellular stimuli (Section 1.10). These sRNAs operate through varied regulatory mechanisms, that involve base-pairing with target RNAs and binding to regulatory proteins (Storz et al., 2011; Waters and Storz, 2009). This class of sRNAs and their roles in enterobacteria will be the focus of this work.

Compared to *cis*-encoded sRNAs, *trans*-encoded sRNAs have limited and imperfect base-pairing interactions with their RNA targets (Fig. 1.4 B). They are short transcripts,

typically 50-300 nt long, that regulate RNA targets encoded at distinct genomic sites (Waters and Storz, 2009).

Trans-encoded sRNAs are synthesized under very specific growth conditions and are a central class of regulators with roles in various cellular processes: stress responses, adaptation to new environments, nutrient utilization, quorum sensing, biofilm formation, virulence and drug resistance in prokaryotes (Papenfort and Vogel., 2014; Mika and Hengge, 2014; Holmquist et al., 2010; Chao and Vogel, 2016; Romby et al., 2006; Gottesman and Storz, 2011; Waters and Storz, 2009, Michaux et al., 2004).

In *E.coli,* many *trans*-encoded base-pairing sRNAs are part of regulons induced under specific stress conditions, such as envelope stress (σ^E-induced MicA, RybB, MicL; Guo et al., 2014; Rhodius et al., 2005; Section 4.1), iron deprivation (RyhB; Benjamin and Massé, 2014; Massé et al., 2005), glucose-phosphate stress (SgrS; Vanderpool, 2007), amino acid availability (GcvB; Sharma et al., 2007), oxidative stress (OxyS; Gonzalez-Flecha and Demple, 1999), glucose availability (Spot 42 - Spf; CyaR; DeLay and Gottesman, 2009; Görke and Vogel; 2008; Beisel and Storz, 2011), low temperature (DsrA; Repoila and Gottesman., 2001), osmotic stress (MicF; RprA; Majdalani et al., 2002) and entry into stationary phase.

Functionally, the limited requirement for base-pairing has pervasive consequences: a *trans*-encoded sRNA can regulate multiple targets, a single transcript can be regulated by several sRNAs and a single sRNA can regulate the same target by more than one mechanism (Brobovskyy and Vanderpool, 2016; Repoila and Darfeuille, 2009; Prévost et al., 2011). The sRNA-mediated regulation of multiple targets is subject to competition between targets, as dictated by their affinity for binding and copy-number, sRNA expression level, and Hfq availability (Hussein and Lim, 2011; Moon and Gottesman, 2011; Wagner, 2013). The regulatory implications of all the above are important: these factors allow the complex integration of sRNAs in regulatory networks, thus cells are able to coordinate responses to multiple input signals (Section 1.9). Moreover, not only some sRNAs directly regulate the expression of

many targets (e.g. GcvB; RyhB; Masse et al., 2005) and some directly regulate TFs or sigma factors (e.g. DsrA, MicL, MicF, MicA; Mika and Hengge, 2014; Melamed et al., 2016; Majdalani et al., 1998), but also the regulatory effects propagate through interconnected pathways (Section 1.9)

1.4. Regulatory mechanisms employed by *trans*-base-pairing sRNAs

Mechanistically, base-pairing sRNAs are similar to the eukaryotic microRNAs (miRNAs; Bartel et al., 2009) in their ability to control target stability and translation efficiency (Wagner and Romby, 2015; Storz et al., 2011; Waters and Storz, 2009). However, sRNAs are a much more heterogeneous group, that differs in origin/biogenesis (Section 1.7) and the sites and details of base-pairing with cognate targets. Bacterial sRNAs are generally not constitutively processed (with a few exceptions; e.g ArcZ and RprA sRNAs; Mandin and Gottesman, 2010, Papenfort et al., 2009, Papenfort et al., 2015; Chao et al., 2017) and base-pair in most cases with the 5'-end of mRNAs, rather than with the 3'-UTRs (Gottesman, 2005). However, sRNAs can base-pair at virtually any feature of a transcript and can employ various mechanisms and non-canonical modes of regulation. This section will expand on the mechanisms employed by sRNAs to exert their regulatory function.

General aspects of sRNA-mRNA base-pairing

The sRNA seed region required for base-pairing is typically conserved in enterobacteria, located in single-stranded regions (Peer and Margalit, 2011), encompasses ~6-12 nucleotides (Papenfort et al., 2010; Balbontin et al., 2010) and in some cases, is sufficient for regulatory function (Papenfort et al., 2010; Pfeiffer et al., 2009). Despite the limited size of the pairing region, the sRNAs specifically recognise their targets. Base-pairing is often mediated by the RNA chaperone proteins Hfq (Fig. 1.4) and ProQ (Smirnov al., 2017) that stabilize the sRNA, and facilitate molecular contacts with the target RNA (Vogel and Luisi, 2011; Kawamoto et al., 2006; Zhang et al., 2002; Moller et al., 2002; Sledjeski et al., 2001; Section 1.5). ProQ may chaperone

RNAs that are defined by a high degree of structure (Smirnov et al., 2017).

Most sRNAs negatively regulate their targets (De Lay et al., 2013), but many examples of positive regulation have been reported (Papenfort and Vanderpool, 2015; Fröhlich et al., 2013; McCullen et al., 2010; Soper et al., 2010; Majdalani et al., 1998). The sRNA-target base pairing modulates mRNA synthesis, stability and/or translational efficiency, resulting in either an increase or decrease in protein synthesis. Notably, an sRNA can have an inhibitory effect on a mRNA, and an activating role on a different mRNA (e.g. RyhB- Prevost et al., 2007). Usually, the same seed sequence within sRNAs is used for target interactions, suggesting that the molecular information determining the regulatory effect (positive or negative) is partially embedded in the target mRNA (Papenfort and Vanderpool, 2015). Mechanistically, there are many variations to the canonical translational inhibition mechanism, that include transcript stabilization, modulation of premature termination, and sRNA sponging (Section 1.6).

Post-transcriptional mechanisms of negative regulation by sRNAs

As negative regulators, sRNAs most frequently inhibit translation (De Lay et al., 2013; Desnoyers et al., 2013). Initial studies suggested that most sRNAs base-pair to target mRNAs at regions at or near the RBS (-20 to +15 nt from start codon, often referred to as the start codon window) to block ribosome binding (Fig. 1.5A) (Altuvia et al. 1998). The region on mRNAs that can be targeted to inhibit translation initiation has been since been extended (Holmqvist et al., 2010; Bouvier et al., 2008; Papenfort et al., 2010; Bobrovskyy and Vanderpool, 2016). New mechanisms of interference with translation, such as interference with translation coupling between adjacent ORFs (Vecerek et al., 2007), targeting of translation enhancer elements (Sharma et al., 2007), repression by binding far upstream in the 5'-UTR of targets (Holmqvist et al., 2010; Papenfort et al., 2010), or repression via targeting translation-activating stem loops within mRNAs coding sequence (Jagodnik et al., 2017) keep being discovered. In most cases, translational repression of a target mRNA is coupled to target degradation (Fig. 1.5A), either caused by reduced ribosome occupancy (Wagner,

2009) or by recruitment of RNase E (Ikeda et al., 2011; Prévost et al., 2011; Aiba, 2007; Morita et al., 2005; Massé et al., 2003). Although target degradation has been often observed upon sRNA expression, translation repression is sufficient to achieve gene silencing whilst degradation renders the repression irreversible (Desnoyers et al., 2013; Morita et al., 2006).

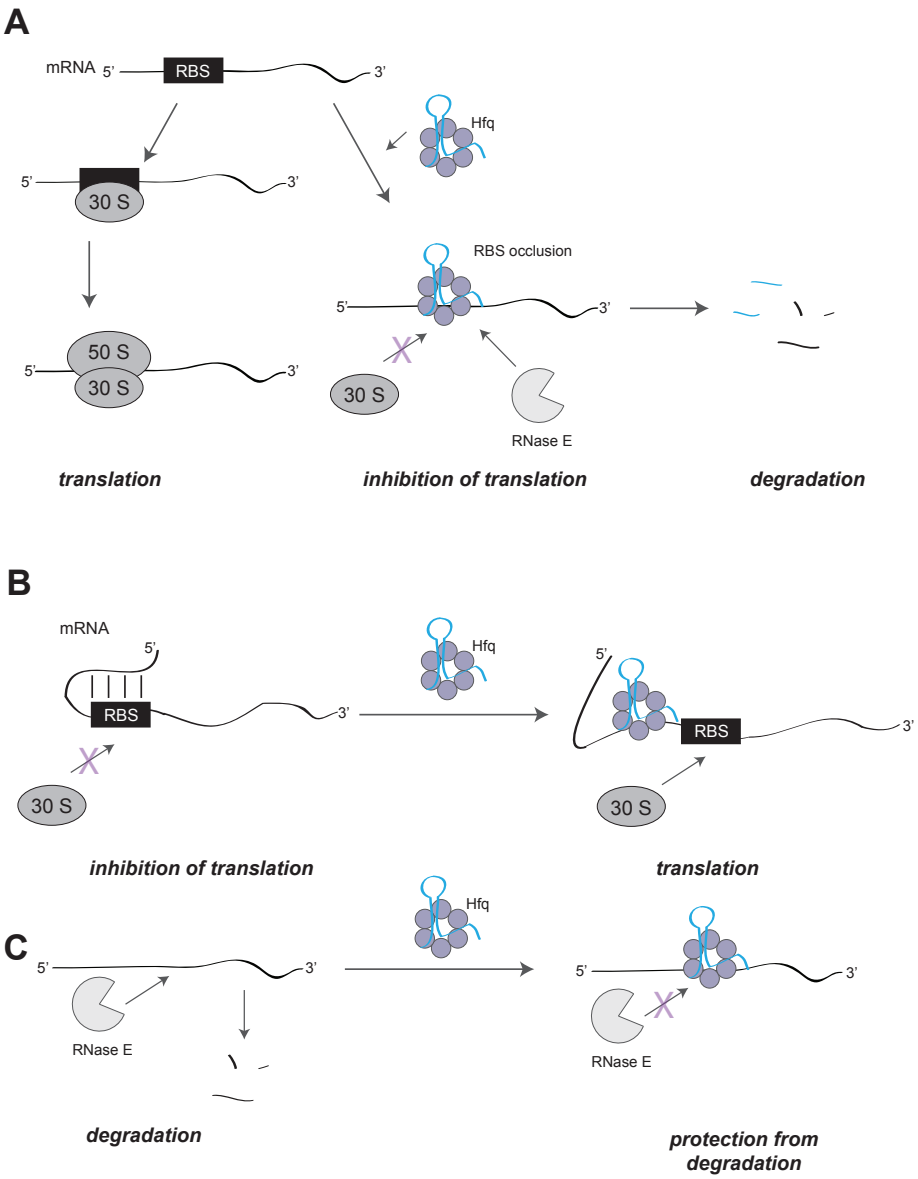

Fig. 1.6: Canonical post-transcriptional mechanisms employed by *trans*-encoded sRNAs in enterobacteria to regulate target expression. (A) An sRNA associated with Hfq may occlude the ribosome-binding site (RBS) of a target mRNA, thus blocking 30S ribosomal subunit binding and inhibiting translation. This mode of repression is often followed by degradation of the target mRNA, or coupled degradation of sRNA and mRNA, usually by RNase E. (B) sRNAs associated with Hfq can activate translation by melting inhibitory secondary structures that hinder ribosome access to the RBS. (C) Hfq associated with an sRNA can also occlude RNase E recognition sites, thus protecting the target RNA from degradation.

Some sRNAs can also actively direct the transcript for cleavage and degradation, without translational inhibition (Papenfort et al., 2010; Pfeiffer et al., 2009), this type can occur at variable locations within mRNAs, including deep within coding sequences or even intergenic regions (RyhB-*iscSUA mRNA*; Desnoyers et al., 2009).

As a consequence of negative regulation, sRNAs and their targets can both be degraded (Overgaard et al., 2009; Masse et al., 2003; Moll et al., 2003) or only one of them is degraded, allowing sRNA recycling. However, these two modes are interchangeable depending on the mRNA target (Overgaard et al., 2009; Figueroa-Bossi et al., 2009), suggesting that the fate of the sRNA is dictated by the details of the sRNA-target duplex, such as secondary structure or identity of their 5'-termini (Section 1.7).

Post-transcriptional mechanisms of positive regulation by sRNAs

In sRNA-mediated positive regulation, pairing increases translation and/or protects target mRNA from degradation (Papenfort and Vanderpool, 2015; McCullent et al., 2010; Fröhlich et al., 2013; Fröhlich and Vogel., 2009; Urban and Vogel., 2008).

To directly activate translation, sRNAs employ an anti-antisense mechanism, whereby they base-pair at the 5'-UTR of target mRNA to disrupt an intrinsic inhibitory structure that sequesters the RBS and the start codon (Fig. 1.5 B) (Prévost et al.,2007; Majdalani et al., 1998). Binding of the sRNA leads to the restructuring of the mRNA, freeing the RBS, which allows translation initiation.

Some sRNAs activate gene expression by blocking RNase E cleavage sites, thereby altering processing and increasing target mRNA stability (Fig. 1.5 C) (Fröhlich et al., 2013; Opdyke et al., 2004).

Transcriptional control exerted by *trans*-encoded sRNAs

By modifying ribosome access or inducing degradation of their targets, sRNAs regulate the coupling between transcription and translation.

However, sRNA activity also directly or indirectly alters transcription initiation, elongation and termination by various mechanisms. Transcription initiation can be regulated by the 6S sRNA that associates with RNAP (Wassarman et al., 2007). Transcription elongation and the co-transcriptional regulation of translation may be controlled by sRNAs as part of ribonucleoprotein complexes that also include RNAP (Van Nues et al., 2015). The 5'-UTRs of transcripts are commonly subject to Rho-dependent termination in *E.coli* (Sedlyarova et al., 2016) and sRNAs have been shown to regulate premature termination either directly, or indirectly.
The first role for sRNAs in the control of transcription termination was established for the indirect transcriptional attenuation by the ChiX sRNA in *Salmonella*: ChiX base-pairs with the leader sequence of its target, *chiP*, repressing translation initiation co-transcriptionally. The resulting uncoupling of transcription from translation promotes Rho-dependent termination of the nascent mRNA (Bossi et al., 2012).

However, the first evidence of direct involvement of sRNAs in Rho-mediated termination was only recently described (Sedlyarova et al., 2016). Genes can be activated by a sRNA-mediated anti-termination mechanism whereby sRNAs anneal to the 5'-UTRs to inhibit premature Rho-dependent termination by sterically blocking Rho translocation along the mRNA strand (Sedlyarova et al., 2016).

Some sRNAs can also act as anti-silencers. A well documented example is DsrA, which counteracts transcriptional silencing of rcsA by the nucleoid-associated H-NS protein (Sledjeski and Gottesman, 1995). DsrA was proposed to prevent H-NS binding to promoters, or to interfere with the silencing activity of H-NS by preventing it from forming a chromatin-like silenced conformation (Sledjeski and Gottesman, 1995).

Examples of sRNAs that modulate the activity of proteins

Other sRNAs act as sponges that sequester RNA-binding proteins (Fig. 1.5). One such protein is CsrA, a global post-transcriptional regulator that binds to specific motifs in the 5'-UTRs of target mRNAs to regulate carbon metabolism, biofilm

formation and motility (Romeo et al., 2013). CsrA consists of a dimer that contains two RNA-binding pockets with high affinity for GGA motifs within loops of short hairpin structures (Schubert et al., 2007; Dubey et al., 2005). The CsrA-mediated regulation is antagonized by the action of two Hfq-independent sRNAs, CsrB and CsrC (Liu et al., 1997; Weilbacher et al., 2003), and one Hfq-associated sRNA, McaS (Jørgensen et al., 2013). CsrB/CsrC and McaS contain the CsrA recognition sequence in multiple copies, thus, by a mimicry mechanism, they act as direct competitors for CsrA target mRNAs (Fig. 1.6 A).Intriguingly, McaS was found to associated with other RBPs, including Hfq, ProQ and RNAP, suggesting that sRNAs can act as scaffolds onto which various proteins can assemble to perform various regulatory roles on the sRNA target (Van Nues et al, 2015).

The 6S sRNA was shown to inhibit transcription by directly binding to σ^{70}-associated RNAP (Fig. 1.6 B). By mimicking the DNA conformation of open promoters, the 6S RNA locates in the active site of RNAP and blocks access to promoters (Wassarman, 2007). This interaction results in the inhibition of transcription from a subset of σ^{70} promoters (Neusser et al, 2008; Cavanagh et al., 2010) and activates transcription

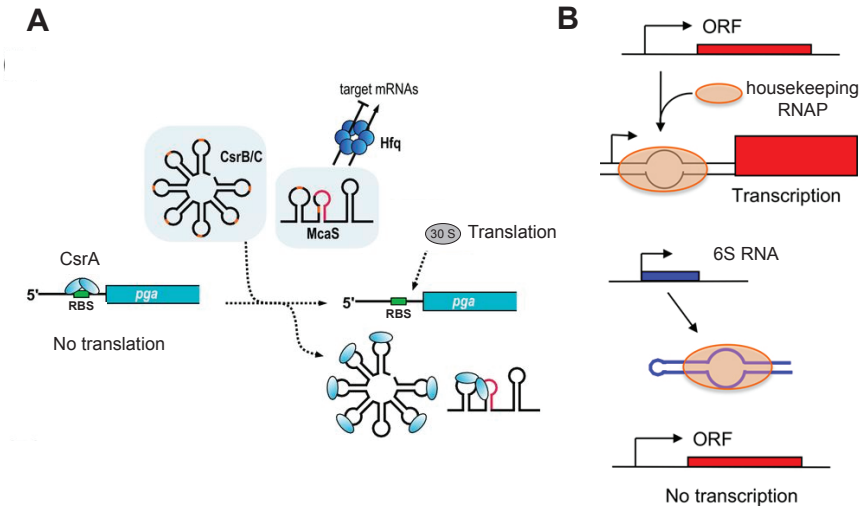

Fig. 1.2: sRNAs modulate the activity of a repressor. (A) CsrA protein downregulates its target genes (e.g. pga) by blocking the SD sequence, thus repressing translation. CsrB/C and McaS bind CsrA to sequester it and derepress its target genes. Addiitionally, McaS also acts as a base-pairing sRNA aided by Hfq, providing a link between the CsrA and Hfq regulons;. Figure adapted from Papenfort and Vanderpool, 2015. (B) 6S sRNA expression can sequester RNAP. Figure adapted from Faucher and Schuman, 2011.

from alternative sigma factor promoters, partly by influencing the competition between the cognate RNAP holoenzymes. Since 6S expression increases during the stationary phase of growth, it affects mainly the balance between the σS (stationary-phase σ factor) and housekeeping RNAP holoenzymes (Wassarman, 2007; Trotochaud and Wassarman, 2005).

In a regulatory cascade contributing to control sugar metabolism, GlmY mimics the structure of another sRNA, GlmZ, to titrate away the adaptor protein YhbJ that is required in RNase E mediated inactivation of GlmZ (Reichenbach et al., 2008; Urban and Vogel, 2008). Consequently, GlmZ is protected from the cleavage of the site that regulates its target, glmS (Gorke and Vogel, 2008; Kalamorz et al., 2007).

Considering the many examples and mechanisms of sRNA-mediated regulation that have been described, it is becoming clear that interaction and competition between sRNAs, mRNAs, RBPs and ribosomes, as well as the folding of RNA species involved all define the efficiency and outcome of the regulation. How many RBPs are involved and how this interplay dictates sRNA-mediated regulation is still not well understood.

1.5. RNA binding proteins with central roles in sRNA-mediated regulation

The flexibility of the cell response to environmental changes requires a rigorous coordination between sRNAs and protein regulators. Apart from the proteins that are targeted/sequestered by sRNAs, a number of RBPs are directly involved in sRNA mediated regulation in Gram-negative bacteria. Most notable are the Hfq protein, RNase E and ProQ (Smirnov et al., 2017; Smirnov et al., 2016). In many Gram-positives or the role of Hfq in sRNA function is accessory (Rochat et al., 2012; Bohn et al., 2007)., but in enterobacteria it is usually required for sRNA regulation.

This section will introduce the main mechanisms employed by the prokaryote degradation machinery and RNA chaperone Hfq to assist sRNAs in their regulatory activity.

Hfq: role in sRNA-mRNA base-paring

In Gram-negative bacteria Hfq plays a major role in *trans*-encoded sRNA-mediated regulation stabilizes sRNAs by protecting them from ribonucleases and facilitates sRNA-target annealing (Gottesman and Storz, 2011; Vogel and Luisi, 2011; Brennan and Link, 2007).

Hfq deletion has pleiotropic effects on Gram-negative bacterial physiology, including general stress response pathways and virulence. In enterobacteria, lack of Hfq alters the expression of approximately a fifth of all genes. Although these dramatic outcomes can be partially caused by indirect effects or other impaired sRNA-independent regulatory functions that require Hfq (Mohanty et al., 2004), they are at least partly attributed to its essential role in sRNA function (Chao and Vogel 2010; Guisbert et al. 2007).

Hfq is a ring-shaped homohexamer with multiple RNA-binding faces: distal, proximal, and lateral (the rim) (Vogel and Luisi, 2011; Sauer, 2013). The proximal and distal faces of Hfq specifically bind sRNA and mRNA targets, respectively, to increase the local concentration of RNA pairs, whilst the rim has been proposed to promote annealing (Panja et al., 2013). There is evidence that there is greater flexibility in the interactions of Hfq with sRNA and mRNAs, and that the mode of binding has regulatory consequences (Schu et al., 2015). Despite the degenerate recognition sequences and heterogeneity sRNA species, how Hfq preferentially binds the pairing sRNAs and their mRNA targets is still unknown. The detailed mechanism by which Hfq employs its chaperone activity to promote sRNA-target regulation also remains unclear. It has been reported that Hfq assists base-pairing by increasing annealing rates (Hwang et al., 2011; Fender et al., 2010), by stabilizing the formed duplex (Soper et al., 2010) or by inducing secondary structure rearrangements to promote the pairing (Maki et al., 2010). The degree of involvement of Hfq in facilitating direct contacts between sRNAs and their targets is likely dependent on the sRNA-mRNA pair (Storz et al., 2011).

The rapid (minute-scale) sRNA-mediated responses *in vivo* do not appear to be in agreement with the measured low dissociation constants and long half-lives of Hfq-RNA complexes (Fender et al., 2010). To solve this paradox, a widely-accepted model proposes that sRNAs and mRNAs actively cycle on and off Hfq (Fig. 1.7), to allow sufficient RNA encounters within short time-frames (Fender et al., 2010). The cycling involves rapid sequential binding and release of RNAs on individual Hfq monomers, that is dependent on the concentration of the RNA species, and greatly affected by competitor RNAs (Wagner., 2013; Fender et al., 2010). The competitor

RNA transiently associates with Hfq-RNA complexes, the RNAs exchange binding sites, until one of them dissociates. This mode of binding allows efficient use of the Hfq pool, which is limiting. Consistent with this model, RNAs have been reported to compete for Hfq, and that an abundant sRNA can indirectly impair the regulation of others (Moon and Gottesman, 2011; Hussein and Lim, 2011). Therefore, sRNA-mediated regulation is greatly dependent on the intracellular levels and availability of sRNAs and their targets, copy-number of interacting proteins, as well as the kinetic parameters of the interactions (Storz et al., 2011). The importance of Hfq RNA exchange on the global effects of sRNA-mediated regulation as well as its impact on patterns of multi-target regulation are increasingly being recognized. For example, cycling of RNAs on Hfq can impact sRNA-mediated multi-target regulation. The

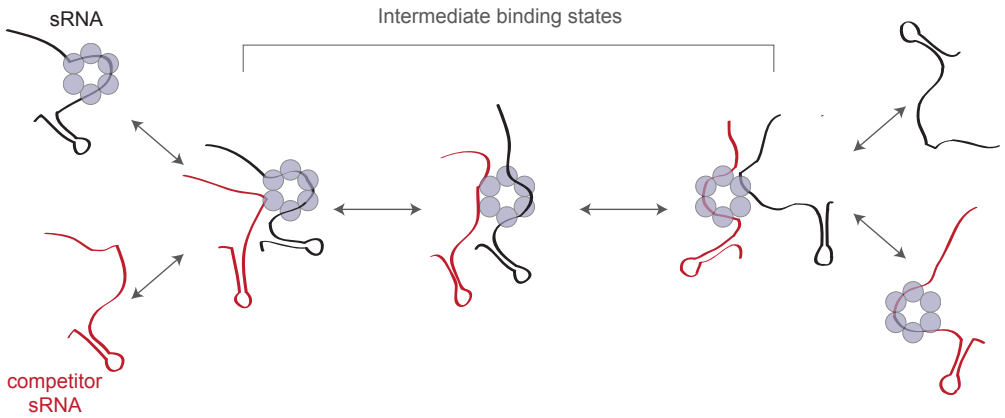

Fig. 1.7: Simplified representation of dynamic sRNA cycling on Hfq: Hfq-bound RNAs are continuously and reversibly (double arrows) displaced by competitor RNAs. Entry of competitor sRNA starts with association of one subunit, followed by dynamic exchange of binding sites across the surface of the hexamer (the intermediate binding states) until the resident RNA is displaced. Figure reproduced/re-drawn from Fender et al., 2010.

hierarchy of target effects is defined by sRNA-target association (Fei et al., 2015), but also the Hfq occupancy of the involved sRNAs. Since the latter is most dependent on the competition for Hfq, the cycling kinetics could affect the global pattern of target RNA expression upon sRNA induction (Wagner, 2013).

RNase E and its intimate links with sRNAs

In *E.coli,* the endonuclease RNase E has a central role in controlling intracellular levels of all transcript classes, as it is responsible for the degradation and/or processing of mRNAs and stable RNAs. Together with polynucleotide phosphorylase (PNPase), RNA helicase B (RhlB), enolase and other proteins (Carpousis, 2007; Bandyra et al., 2013) it forms a complex that is referred to as the RNA degradosome. RNase E is responsible for the initial endonucleolytic cleavage that triggers complete degradation of most mRNAs (Clarke et al., 2014). RNase E is also involved in the turnover of sRNAs that are not bound by Hfq and sRNA-target interactions mediated by Hfq (Bandyra et al., 2012; Moll et al., 2003; Masse et al., 2003).

Moreover, many RNAs are synthesised as precursors that must be processed to their mature forms to be functional – and the decay of mRNAs and and RNA maturation have overlapping steps (Deutscher, 2006). RNase E is essential in the maturation of stable RNAs: tRNAs (Liz and Deutscher, 2002) and rRNAs (Li et al., 1999), but is also involved in the processing of sRNAs and mRNAs to yield shorter, stable species (Section 1.7). RNase E has little sequence specificity, but until recently was broadly defined as having a preference for single-stranded A/U-rich regions (Carpousis, 2007). A RNase E cleavage signature has only recently been more precisely defined: recognition sites tend to be located in single-standed regions, two nucleotides upstream of a U ribonucleotide (Chao et al., 2017).

The endonuclease activity of RNase E is strongly potentiated by accessible 5’-monophosphate termini (5’-P) on the substrate (Jiang and Belasco, 2004). Thus most newly transcribed RNAs bearing 5’-triphosphate groups are not targeted by

RNase E. The sRNAs, however, mostly have a single 5'-phosphate group. Therefore, by base-pairing with their targets, sRNAs can provide "specificity" by guiding RNase E to its cleavage site within the sRNA target. (Bandyra et al., 2012).

RNase E not only contributes to the degradation of sRNAs and their targets (Fig.1.5A) but also assists sRNA-mediated regulation by processing a few sRNAs to yield an active form. For example, endonucleolytic cleavage of the ArcZ sRNA exposes the seed sequence (Papenfort et al., 2009). RNase E is also required for processing of the 6S sRNA, which involves both endo- and exonucleolytic activity: the 5' processing of precursor ssrS is mediated by RNase E, whereas for the exoribonucleolytic trimming for 3' processing RNases T and PH are most important (Zhongwei et al., 1998, Chae et al., 2011). Alternatively, RNase E cleavage of mRNAs has also been shown to generate sRNAs that originate from the UTRs (Section 1.7).

Although RNase E is a key player in the sRNA-mediated post-transcriptional regulation, other RNAses are also involved (Guo et al., 2014; Opdyke et al., 2011). How the other RNases and RBPs interact with sRNAs and modulate their activity is still poorly understood.

1.6. Regulating the regulator: bacterial RNA sponges antagonize sRNAs

RNA sponges are regulatory species that directly control the activity of other riboregulators, and have been described in eukaryotes as antagonists of miRNAs (Salmena et al., 2011; Hansen et al., 2013). In prokaryotes RNA sponges generally antagonize the activity of sRNAs (Azam and Vanderpool, 2015; Myiakoshi et al., 2015; Lalaouna et al., 2015; Tree et al., 2014; Figueroa-Bossi et al., 2009) by base-pairing, followed by repression of its activity. RNA sponges can constitute virtually any RNA class: mRNAs, tRNAs or non-coding RNAs.

The first described RNA sponge in *E.coli* and *Salmonella* was shown to be derived from a mRNA and to control the activity of ChiX, an abundant sRNA (Figueroa-Bossi

et al., 2009) (Fig.1.8 A). In the absence of chitosugars, the ChiX sRNA represses the expression of *chiP,* encoding the chitoporin responsible for chitosugar uptake. In the presence of chitosugars, the *chb* mRNA encoding products for chitosugar metabolism is processed, generating a sponge that base-pairs with ChiX and destabilizes it.

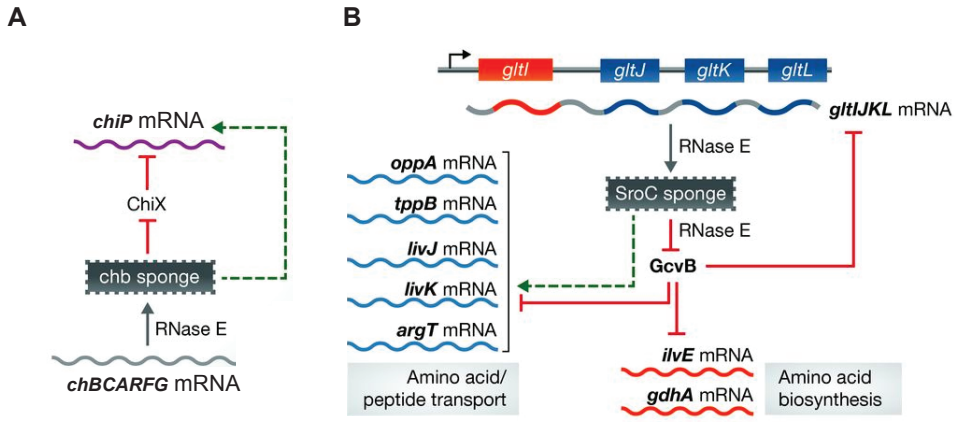

Fig. 1.8: Examples of sRNA sponges in bacteria that promote mRNA crosstalk. Cross-talk is indicated by dashed arrows. (A) The chb sponge originates from the chBCARFG parental RNA degrades ChiX, allowing cross-talk between chitosugar transport and metabolism genes. (B) The sponge RNA SroC, processed from a GcvB target, base-pairs with GcvB and degrades it, causing derepression of its targetome, including the precursor mRNA of SroC. SroC mediates cross-talk between GcvB targets. Image adapted from (Azam and Vanderopool, 2015)

Reduced levels of ChiX result in derepression of *chiP* and net increase in chitoporin synthesis (Plumbridge et al., 2014; Figueroa-Bossi et al., 2009; Overgaard et al.,2009). Similarly, the RprA sRNA is sponged by *csgD* mRNA (Mika et al., 2012). The 3′ -UTR region of *pspG* acts as a sponge of the Spf sRNA (Melamed et al., 2016).

Sponges derived from tRNA precursors control the activity of RybB and RyhB sRNAs, and possibly also MicF (Lalaouna et al., 2015). These sponges are stable intermediates of tRNA processing, originating from the internal transcribed spacers (ITS) and external transcribed spacers (ETS) of tRNA operons. For instance, the 3′ ETS of the glyW-cysT-leuZ base-pairs with RybB and RyhB sRNAs, reducing their effective concentration. Lalaouna et al proposed that this activity is necessary to adjust the sRNA activity to the growth conditions. Specifically, under non-inducing conditions, the transcriptional repression by Fur is not sufficient to abolish RybB

synthesis. 3′ETS^{leuZ} , consequently, could act as a sponge on the excess sRNA to adsorb transcriptional noise in RyhB expression (Lalaouna et al., 2015).

In enterohemorrhagic *E.coli*, bacteriophage-encoded sRNAs (e.g. AgvB) have been reported to act as sponges that inhibit sRNA activity (Tree et al, 2014). Interestingly, AgvB targets the GcvB sRNA, but does not facilitate its degradation.
SroC was the first bacterial sponge in which a sponge-mediated cross-talk was demonstrated to affect multiple mRNAs within a single regulon (Myiakoshi et al., 2015b). SroC originates from the polycistronic *gltIJKL* mRNA encoding the glutamate/ aspartate ABC transporter. GcvB sRNA inhibits translation initiation of the *gltIJKL* mRNA by base-paring to the 5'-UTR of *gltI* (Sharma et al., 2007). However, a region between *gltI* and *gltJ* produces SroC that base-pairs with GcvB and promotes its degradation. Destabilization of GcvB causes derepression of its targetome, including the parental mRNA of SroC. Interestingly, in this case, the sRNA is regulated by a sponge released from one of its cognate mRNA targets. Physiologically, SroC enables activation of many transcripts of the amino acid transport and metabolism pathway, and promtes bacterial growth when peptides are the sole carbon and nitrogen source.

SroC was later shown to directly interact with and down-regulate the levels of another sRNA, MgrR, relieving the MgrR-mediated repression of eptB, encoding a lipo-polysaccharide (LPS)-modifying enzyme (Acuna et al., 2016). SroC base-pairs with MgrR in the region encompassing the seed sequence for interaction with the *eptB* mRNA and antagonizes its regulatory effects on gene expression and LPS modification.

To date, only a number of RNAs that act as sponges have been described in bacteria, however, it is becoming clear that they are an emerging group. For instance, RNase E CLASH (Section 1.8) performed in enterohemorrhagic *E.coli* uncovered ~150 unique sRNA-sRNA interactions (Waters et al., 2017), suggesting that this mode of control of sRNA activity is widespread and that more sponges are yet to be discovered. Interestingly, all sponges characterized thus far employ the "repression of a repressor"

mechanism, resulting in the activation of the target genes of the sRNA regulator (Azam and Vanderpool, 2015) and most originate by processing from precursors. The efficiency by which the sponge controls expression of sRNA and its targets depends on a number of parameters, including the relative expression of the sRNA, target mRNA and sponge and the binding affinity intrinsic to each interaction (Myiakoshi et al., 2015b). In conclusion, sponge RNAs provide an additional level of control that allows cells to exploit the cross-talk between physically unlinked genes and to fine tune post-transcriptional circuitry.

1.7. Features and biogenesis of base-pairing sRNAs

A full understanding of sRNA-mediated networks necessitates knowledge of their number and identity. Additionally, to unravel sRNA modes of action, details of their abundance, structure, biogenesis and turnover are required.

Features of typical *trans*-encoded sRNAs

sRNAs vary significantly in length (50-300 nt), sequence and secondary structure (Storz et al., 2011). Typical functional sRNAs have a modular structure (Fig. 1.9). that encompasses at least one seed sequence for base-pairing with targets (Guillier and Gottesman, 2008; Papenfort et al., 2010; Balbontin et al., 2010; Pfeiffer et al., 2009), a binding site for Hfq and a Rho-independent terminator that consists of a GG-rich stem-loop followed by a poly-U track (Otaka et al., 2011).

Due to the lack of defining common features, many sRNAs are likely to remain undiscovered in screens that search for classical properties. Transcriptome-wide approaches, including bioinformatic predictions, microarrays and deep-sequencing experiments led to the identification of hundreds of candidate sRNAs (Argaman et al, 2001; Sharma and Vogel., 2009; Wassarman et al., 2001; Vogel et al, 2003; Raghavan et al, 2011). Computational predictions are conventionally based on sequence conservation, structural homology, and presence of common features such as Rho-

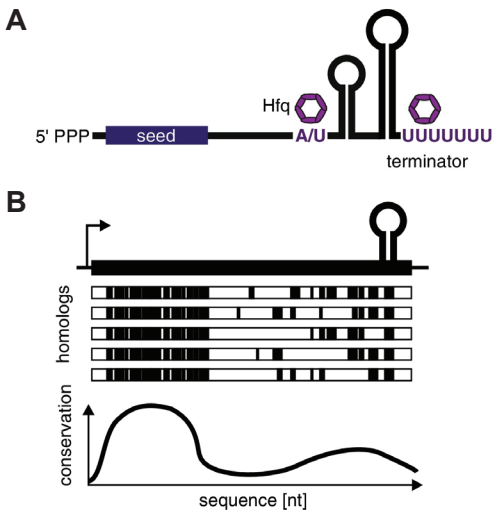

Fig. 1.9: Typical features of Hfq-associated sRNAs (A) Modular structure of sRNAs: Most sRNAs are synthesisd as primary transcripts with a 5'-PPP group at their 5'-end. and contain a seed sequence for base-pairing with target mRNAs (blue), Hfq-binding regions and a Rho-independent terminator. Hfq-binding sites are near the Rho-independent terminator regions. (B) A schematic representation that shows that typically, functional sRNAs are well-conserved and that the seed sequence of a sRNA is typically most conserved region among homologs. Image adapted from Myiakoshi et al,, 2015a.

independent terminators and independent promoters (Peer and Margalit, 2011). Proven interaction with Hfq in transcriptome-wide studies (Zhang et al, 2003; Sitka et al, 2008; Chao et al., 2012; Tree et al., 2014) may further confirm whether a given ORF is a functional sRNA. However, most studies looked for sRNAs in unannotated genomic regions, particularly intergenic regions (IGRs).

3'-UTRs of transcripts as a source of sRNAs

Some of the key features of sRNAs are also shared by some protein-coding RNAs, namely sequence conservation and presence of Rho-independent terminators (Sauer et al., 2011). Intriguingly, recent studies have identified stable 3'-UTR fragments of mRNAs that had the same functional and structural features of sRNAs (Vogel et al., 2003; Zhang et al., 2003). Because of their proximity to mRNA reading frames these sRNA candidates probably remained undetected in standard IGR-based screens.

These 3'UTR fragments were also enriched in Hfq pull-down experiments (Chao et al., 2012; Tree et al., 2014), suggesting they could represent functional Hfq-associated sRNAs.

Based on their biogenesis, 3'-UTR overlapping sRNAs were classified as either type I or type II (Chao et al., 2012; Myiakoshi et al., 2015a). Type I sRNAs (Fig. 1.10 A) are transcribed from independent promoters within mRNAs. Notable examples are DapZ (partially overlapping dapB mRNA; Chao et al., 2012) and MicL (overlapping cutC; Guo et al., 2014).

Type II sRNAs lack an independent promoter and originate by processing of the parental mRNA (Fig. 1.10 B). Thus, in this case the sRNA level is strictly dependent on the synthesis of the parent RNA (Chao et al., 2012). In most cases, RNase E is the endoribonuclease responsible for releasing the sRNA (Mackie et al., 2013; Clarke et al., 2014). An early discovered example of type II sRNA is DicF, derived from a polycistronic prophage mRNA that encodes a protein involved in cell-division. The importance of type II sRNAs has increased with the characterization of sRNAs with established roles in bacterial physiology. SroC is released by RNase E-mediated

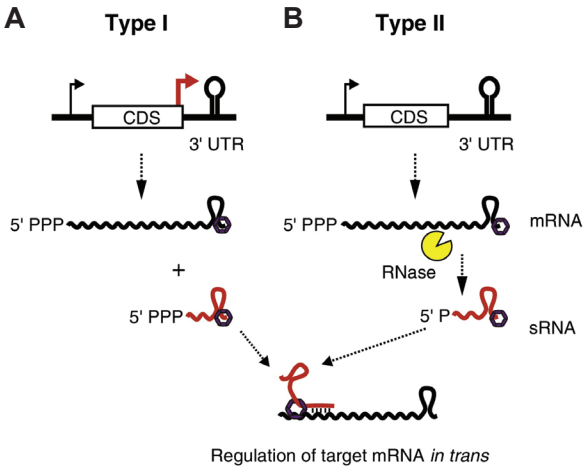

Fig. 1.10: Alternative sRNA biogenesis pathways from the 3'-UTRs of mRNAs. (A) A sRNA can be independently transcribed from an internal promoter (red) within the mRNA gene, and bears a characteristic 5'-PPP end. (B) sRNAs can be processed by ribonucleases (yellow) from parental mRNAs, and bear a 5'-P ends; CDS indicates mRNA coding sequence. Image adapted from Myiakoshi et al,, 2015a.

processing of the first cistron of the *gltIJKL* polycistron and regulates amino acid pathways. CpxQ is similarly cleaved from the 3'-UTR of *cpxP* mRNA and is involved in envelope stress responses (Chao and Vogel, 2016). GadF is processed from the 3'-UTR of gadE and regulates acid-stress responses (Melamed et al., 2016). A ProQ-associated sRNA, RaiZ that represses *in trans* the mRNA of histone-like protein HUα, is also released by processing (Smirnov et al., 2017).

Type I sRNAs have a 5′-triphosphate (5′ -PPP), whereas type II sRNAs possess a 5'-mono phosphate (5'-P) as a result of endonucleolytic cleavage of the parent RNAs (Chao et al., 2012; Myiakoshi et al., 2015a). As mentioned above, 5'-mono phosphates provide a regulatory advantage as they promote target cleavage by RNase E (Bandyra et al., 2012). The propensity for RNase E-mediated degradation may also explain why 5'-P bearing sRNAs are not readily detected *in vivo*.

Global studies, such as mapping of 5'-processed transcripts by 5'-end dependent exonuclease treatment followed by RNA-seq (Thomason et al., 2015), RILseq (Melamed et al., 2016), RNase E CLASH (Waters et al.,2017) and TIER-seq (Chao et al., 2017) not only revealed a large number of 3'-UTR-derived candidate sRNAs, but also identified other transcript classes that can serve as resevoirs for sRNAs (Lalaouna et al., 2016). This is not surprising, given that RNase E not only degrades mRNAs (Hui et al., 2014), but is also involved in rRNA and tRNA processing (Apirion, 1978; Kime et al., 2014). Collectively, these observations strongly argue that RNA processing to yield regulatory molecules is a prevalent event.

The detailed characterization of more sRNAs from 3' regions of transcripts is bound to reveal more non-canonical mechanisms of sRNA regulation. For example, most RNA sponges that antagonize sRNAs are generated by processing/degradation (Lalaouna et al., 2015; Myiakoshi et al., 2015b), and the specific features of this type of sRNAs may be correlated with specific regulatory mechanisms.

1.8. Unravelling the roles of sRNAs: targetome-identification based approaches

To date, over 200 sRNAs have been identified in *E. coli* and *Salmonella*, but most of these regulatory molecules have unassigned functions. Discerning sRNA-target interactions is essential for understanding the roles of sRNAs in the cellular networks. To gain a global understanding of the sRNA-mediated reshaping of gene expression, the field has embarked in a quest to unravel the sRNA-target interaction networks.

Two recent technological developments have contributed significantly to our current systems-level knowledge of sRNA regulatory networks: RNA immunoprecipitation and CRAC/CLASH. Both are discussed below.

Transcriptome-wide maps of Hfq binding

RNA immunoprecipitation studies using antibodies against Hfq (RIP-seq) have produced the first maps of Hfq binding across the transcriptome in *E.coli* and *Salmonella* (Chao et al, 2012; Zhang et al., 2013; Bilusic et al., 2014). These studies indicated interactions of Hfq with hundreds of sRNAs and thousands of mRNAs, identified new sRNAs and provided insights into how Hfq facilitates sRNA-mRNA interactions. These methodologies have been improved by incorporation of UV cross-linking (CLIP-seq; Holmquist et al., 2016), enabling the identification of direct RNA targets and increased stringency in the purification of Hfq-RNA complexes (Hfq CRAC; Tree et al., 2014). These studies defined more precisely the patterns of Hfq-RNA target recognition, and enabled improved predictions of sRNA-target interactions (Holmquist et al., 2016).

Transcriptome-wide computational prediction of sRNA targets

The targets of a sRNA can be predicted computationally by sequence based analyses – however, the degenerate patterns of base-pairing and lack of conservation of primary or secondary structures of targets have made accurate predictions challenging. To increase the stringency of *in silico* prediction algorithms, the requirement for a minimum

number of contiguous nucleotides to be involved in the interactions and predicted accessibility of the seed sequences have been incorporated (Peer and Margalit, 2011). Some of the most performant global sRNA-target prediction programs have included phylogenetic conservation of mRNA seed sequences to increase likelihood of discovering functional targets (e.g. CopraRNA; Wright et al., 2013). Even with these improvements, computational prediction methods are prone to both false-positives and false-negatives. Noteworthy, these methodologies should be revised as more regulatory mechanisms of sRNA action are being described. An obvious example is the discovery of sRNA sponges: available prediction tools report only the mRNA class of transcripts as potential targets, and all sRNAs and tRNAs are by default filtered out. Additionally, the sRNA-sRNA duplexes usually encompass short interacting sequences that may not be considered significant by current algorithms. Therefore, parameter adjustment in such screens becomes paramount and more challenging than ever, and must be accompanied by laborious benchmarking.

In vivo capture of sRNA-RNA interactomes

Given these caveats, recently developed *in vitro* and *in vivo* methods aimed to capture sRNA-target interactions. To capture the targets of individual sRNAs, an aptamer tagged sRNA can be expressed *in vivo* and used as a bait to purify interacting RNAs, which are subsequently sequenced (Imig et al., 2015; Lalaouna et al., 2015). This approach, referred to as MAPS, was the first to detect interactions between sRNAs and spacer regions from tRNA precursors (Section 1.6; Lalaouna et al., 2015). With GRIL-seq an RNA ligase is expressed to ligate sRNA-mRNA duplexes *in vivo* to improve the capture of shorter duplexes (Han et al., 2016). Although these approaches contributed to the identification of sRNA target suites for a number of sRNAs, thus enriching sRNA networks, they are limited to the investigation of a single sRNA at a time.

To obtain a truly global snapshot of sRNA networks *in vivo*, two recently published studies experimentally profiled transcriptome-wide sRNA-mRNA interactions in bacteria: RIl-seq (Melamed et al., 2016) and CLASH (Waters et al., 2017) (Fig. 1.11

A). Both rely on proximity-based ligation of base-paired RNAs (Helwak et al., 2013), but differ in the experimental and computational approaches used for the detection of interacting RNAs (Hör and Vogel., 2017). Both approaches used Hfq or RNase

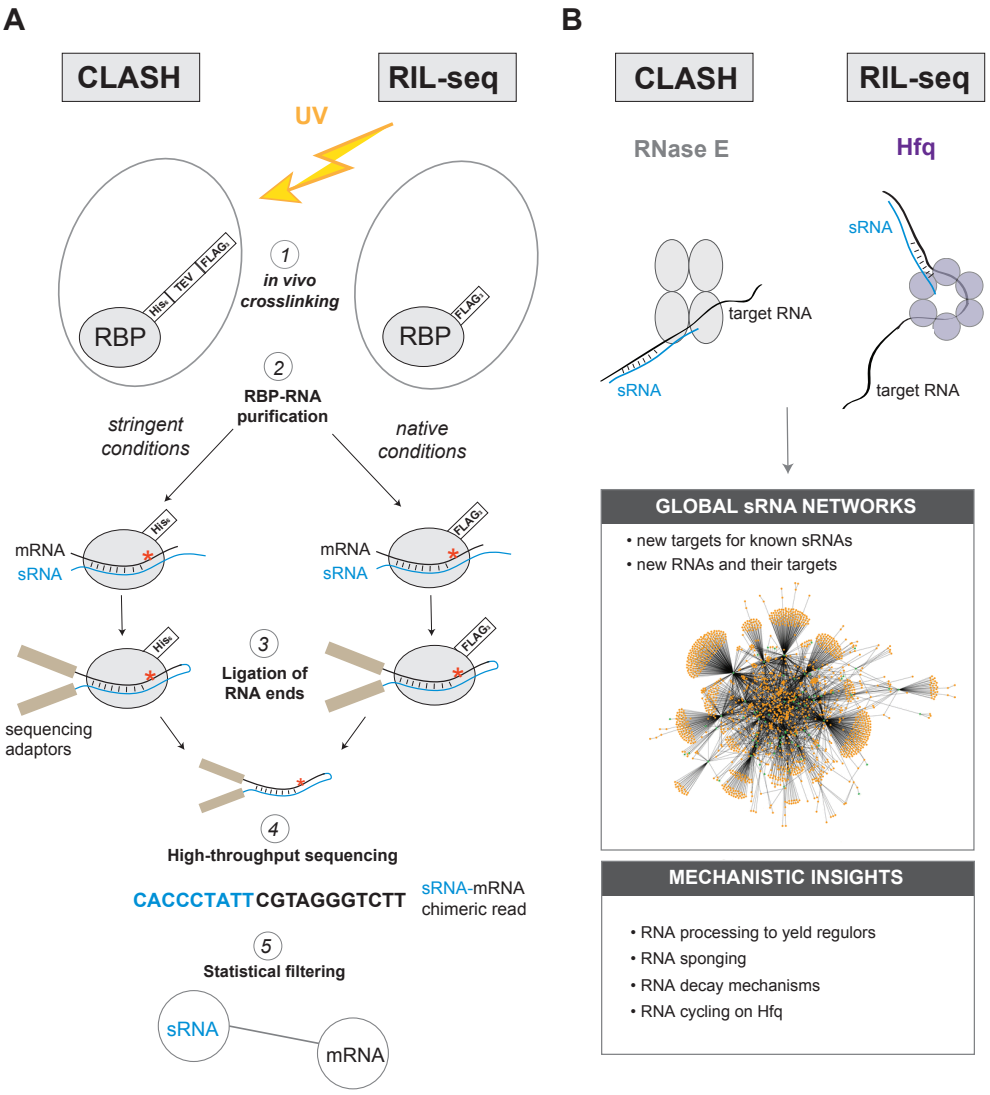

Fig. 1.11: Overview and brief comparison of CLASH and RIL-seq methodologies for capturing RNA-RNA interactions. (A) The critical steps of each method are numbered. Both methods employ *in vivo* UV crosslinking to stabilize RBP-RNA interactions. In CLASH, the RBP used as a bait for capturing RNA-RNA interactions is dual-affinity tagged (His6-TEV-3XFLAG) to allow stringent purification of RBP-RNA complexes. If the RBP binds RNA duplexes, these can be ligated into a hybrid RNA molecule using T4 RNA ligase. In RIL-seq, the RBP of interest is FLAG-tagged and purified under non-stringent conditions, and the RNA ends of RNAs in duplexes are ligated together. In both protocols, the high-throughput sequencing of hybrid RNAs yields chimeric reads that can be mapped to sRNA-target interactions using pipelines and statistical filters specific to each method. (B) The type of information that can be obtained by mapping intermolecular RNA interactions *in vivo* by using either method. CLASH was applied in enterhemorragic *E.coli* using RNase E as a bait, while RIL-seq was applied in *E.coli* using Hfq as a bait, to capture sRNA-target RNA interactions. The network shown represents RNA-target interactions uncovered by RIL-seq (Melamed et al., 2016) and is reproduced from Nitzan et al., 2017.

E as baits, proteins central to sRNA-mediated regulation, for isolating sRNA-target duplexes. These technologies yield single (95-99%) and chimeric reads (1-5%) that contain the sequences of the base-paired RNA fragments that were ligated to each other (Fig.1.11) (Kudla et al., 2011; Waters et al., 2017; Melamed et al., 2016).

The CLASH (UV-crosslinking, ligation and sequencing of hybrids) methodology is based on the CRAC protocol (Granneman et al, 2009). In CRAC (cross-linking and analysis of cDNAs), an improved CLIP method, the target protein has a tandem affinity tag that allows higher specificity purification of RBP-RNA complexes using stringent conditions (Granneman et al, 2009). CLASH originally was basically a bioinformatics data analysis pipeline that allowed direct identification of RNA-RNA duplexes in CRAC data, which are occasionally recovered by chance (Kudla et al., 2011). Later versions of CLASH incorporated new experimental steps, including a separate RNA-RNA ligation step, to enhance the generation and recovery of chimeric reads (Helwak et al., 2013). For the first time, RNA-RNA interactions could be detected directly, by-passing computational predictions (Helwak and Tollervey, 2014). CLASH was employed to characterise RNA-RNA interactions in yeast, and the human miRNA interactome (Kudla et al., 2011; Helwak et al., 2013). Waters et al. implemented this method in *E. coli* to capture RNA-RNA duplexes bound to RNase E (Waters et al., 2017).

Briefly, CLASH involves tagging the protein of interest with a dual affinity tag (HTF: HIS6-TEV-3xFLAG or HTTP: HIS6-TEV-2x PROTA). The cells are then UV-irradiated to forge covalent bonds (“crosslink”) between the bait protein and the bound RNA. Subsequently, the protein of interest is purified under very stringent and denaturing conditions to ensure that only RNAs that are covalently attached to the proteins are purified (Fig. 1.11 A). A fraction of the RNA molecules covalently bound to the protein represent RNA-duplexes, which are then joined together by RNA ligation into a single chimeric RNA fragment (Kudla et al., 2011; Helwak et al., 2013). The chimeric reads are then extracted from the sequencing data using established bioinformatics approaches (Travis et al., 2014). *In silico* folding of the fragments of the chimeric reads makes it possible to determine how the two RNA fragments form a duplex.

Performing CLASH in bacteria is certainly not trivial: An initial attempt to apply CLASH in enterohemorrhagic *E.coli* using HTF-tagged Hfq as a bait yielded an unexpectedly low recovery of chimeric reads (Tree et al., 2014; Waters et al., 2017; Hor and Vogel., 2017). Switching the bait to RNase E led to a much higher recovery of hybrid reads (Waters et al., 2017).

RIL-seq is in principle similar to CLASH (Fig. 1.11 A), but uses a single epitope tag (such as FLAG) to purify Hfq and associated sRNA-RNA duplexes, and non-denaturing or semi-denaturing conditions to purify these duplexes (Melamed et al., 2016). The authors postulate that non-stringent purification conditions preserve the integrity of the Hfq hexamer, thus maintaining it as a scaffold onto which RNA-RNA interactions are also better preserved during capture. Therefore, compared to CLASH, RIL-seq should also enrich weaker and transient interactions. However, since Hfq is associated with many RNAs *in vivo* (e.g. rRNA, tRNA) even non-specifically, low stringency purification implies a higher risk of capturing spurious RNA-RNA interactions. To tackle this, the accompanying bioinformatics pipeline adds stringency to the method by filtering chimeric reads for statistical enrichment (Melamed et al., 2016).

Interestingly, despite their major differences, RIL-seq and CLASH recover a roughly similar count of sRNA-mRNA hybrids in comparable growth conditions (Waters et al., 2017). For both datasets, a significant proportion of the hybrids overlap the seeds of known sRNA-target pairs, thus they are likely to represent true sRNA-target interactions. Known sRNA-mRNA interactions were recovered by both methods – for RIL-seq the overlap was 56 % for all growth conditions studied (Hör and Vogel., 2017). Moreover, similar sRNA seed regions and motifs were identified for abundant, well described sRNAs (e.g. ArcZ, MgrR, GcvB, CyaR) (Hor and Vogel., 2017). Corroborated and accompanied by several experimental validations, these observations indicate that both methods can reliably detect sRNA-RNA interactions *in vivo*. Each dataset has significantly expanded the established networks for the conditions and organisms they were applied in, and revealed major rewiring of networks between growth conditions (Melamed et al., 2016).

Besides providing sRNA interactome maps in *E. coli* (Fig. 1.11 B), RIL-seq and CLASH provided key mechanistic insights into sRNA and RNase E mechanisms of action (Waters et al., 2017) and sRNA-mRNA pairs cycling on Hfq (Melamed et al., 2016). Notably, both methods challenge established views of sRNA-mediated regulation. First, the recovery of sRNA-sRNA interactions suggests that sRNA sponges are more prevalent than previously anticipated (Waters et al., 2017; Melamed et al., 2016). The recovery of hybrids between 3'-UTRs and 5'-UTRs of distinct mRNAs indicates more 3'-UTRs may contain novel sRNAs (Melamed et al., 2016). Moreover, cis-encoded sRNAs can also have *trans*-encoded targets (e.g. GadY, ArrS; Melamed et al., 2016).

Both studies indicate that the classical view that Hfq-dependent sRNAs mainly act in the start codon window of their mRNA targets to inhibit translation may have more exceptions than previously considered, as hundreds of mRNAs appeared to be targeted outside this region (Hör and Vogel., 2007). RIL-seq and CLASH using a single protein as a bait do not have the power to establish on their own whether sRNAs negatively or positively regulate their targets. However, comparison of Hfq CRAC data (Tree et al., 2014) with RNase E CLASH data (Waters et al., 2017) revealed coincidences of RNase E and Hfq sites that support a model whereby RNase E is recruited immediately downstream of the Hfq binding site and cleaves the target, triggering its degradation (Waters et al., 2017). Similarly, comparison of the RIL-seq data with multiple available transcriptome and ribosome profiling datasets allowed the authors to make inferences on the consequences of sRNA-mRNA base-pairing, the strength of the interactions and their association with Hfq (Melamed et al., 2016). An even more recent report (Liu et al., 2017) employed a modified CLASH protocol that aimed to remove the requirement for RBPs as baits for RNA-RNA duplex purification. Instead, the method employs aminomethyltrioxsalen treatment coupled irradiation at 365 nm to generate inter-strands adducts between juxtaposed uridine bases, to lock the interactions. After statistical filtering, only 29 transcript pairs were significant interactions, and only a few of these were sRNA-mRNA interactions. Conceptually, this approach has potential, however, the authors acknowledge that there are important technical problems to be overcome (Liu et al., 2017).

The application of RIL-seq and RNAse E CLASH in *E. coli* constitutes a major advance in unravelling sRNA networks, as they are the first studies that simultaneously capture the Hfq-mediated interactome for all sRNAs. Although, globally, both methodologies revealed common trends, a thorough comparison between the two datasets was not performed. Additionally, in either dataset, due to biases (e.g. ligation biases, spurious RNA-RNA interactions, intrinsic nature of RNA-RNA base-pairing), true biological interactions can still be over- or underrepresented, in spite of careful experimental considerations and statistical filtering (Melamed et al., 2016). This leaves the open question whether the frequency of recovered hybrids for each sRNA or mRNA is truly reflective of their regulatory activity, which can be at least partially addressed by optimizing the quantitative robustness of such methodologies.

1.9. Unravelling the roles of sRNAs in regulatory networks

While regulators and targets for many base-pairs have been identified and validated, and an outstanding number of interactions have been recently uncovered (Melamed et al., 2016; Waters et al., 2017), little is known about how sRNAs participate in these regulatory networks to help cells respond to environmental changes.

Network motifs have been extensively described for transcriptional networks, but given the numerous roles in metabolic and stress response pathways and the multifaceted nature of sRNA-mediated regulation, we need to incorporate sRNA post-transcriptional networks into regulatory circuits to gain a complete understanding of gene expression changes during adaptive responses. Indeed, the importance of sRNAs in regulatory networks has been increasingly recognized (Nitzan et al., 2017; Beisel and Storz, 2010).

sRNAs in network motifs

TFs and sRNAs control target genes and affect regulatory networks both locally and globally by employing regulatory circuits - defined connected patterns of regulators and

targets. Network motifs are circuits that are recurrent and widely utilized in networks (Milo et al., 2002). Due to the differences in TF-based regulation and sRNAs-based regulation, it is advantageous for the cell to integrate both in mixed network motifs to ensure coordinated and rapidly fine-tune the kinetics of responses (Section 1.10). Some of the most common mixed network motifs are the single input module (SIM), dense overlapping regulon (DOR), feedforward loop (FFL) and feedback loop. SIMs have a minimal architecture, in which a single regulator regulates the expression of multiple genes. Most sRNAs in SIMs repress multiple genes in response to changes in the environment (Fig. 1.12 A). By employing SIMs, cells can coordinate positive or negative regulation of multiple genes and introduce hierarchical ordering within networks, presumably dictated by how favourable a regulator-target interaction is (Beisel and Storz, 2010; Levine et al., 2007; Section 1.10). Either the TF or sRNA can be at the top layer. A canonical example of such circuitry is the Fur-repressed RyhB regulon. RyhB becomes activated by low-iron conditions and downregulates many transcripts coding for iron-utilizing proteins (Masse et al., 2005). A DOR combines multiple overlapping SIM - each responding to a different signal (Beisel and Storz, 2010), thus DORs are particularly suited for integration of simultaneous multiple signals. Simple DORs can involve circuitry from multiple sRNAs to a transcriptional regulator. In more complex DORs multiple TFs that regulate multiple sRNAs affect overlapping targets, as exemplified by outer membrane porin (OMP) regulation (Fig. 1.12 B) (Vogel., 2009).

Mixed FFLs and feedback loops: more complex integration of sRNAs in regulatory circuits

Understanding the architecture of FFLs is important, as they are essential building blocks of complex networks (Milo et al., 2002). Given three genes, A, B and C, a feedforward loop (FFL) is a network motif where A regulates expression of B and A and B co-regulate expression of C (Alon, 2007). Each FFL has therefore two regulatory arms for the regulation of C, namely, a direct arm, where A regulates C and an indirect arm, where A regulates C via B. A coherent FFL is defined as a FFL

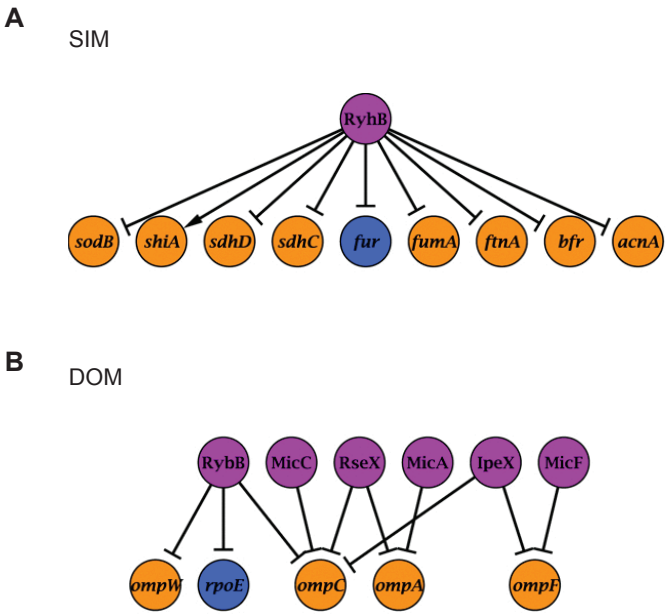

Fig. 1.12: Examples of single-input module (SIM) and dense-overlapping module (DOM) involving sRNAs in enterobacteria. Nodes in pink indicate sRNAs, nodes in blue indicate transcription regulators, nodes in orange indicate target genes (A) RybB downregulates the expression of many genes involved in iron metabolism in response to iron depletion. (B) Major porin expression regulation is tightly regulated in response to various environmental conditions by many sRNAs: one sRNA can downregulate multiple porins and non-porin targets and the same target can be downregulated by multiple sRNAs. Image adapted from Shimoni et al., 2007.

in which both arms work in concert/ together to achieve a regulatory effect (e.g. they are both activating C). In an incoherent FFL, both arms work in opposition to activate or repress C (each arm leads to a different outcome). Since each of the three edges of FFLs can represent either activation or repression, there are eight combinations defining FFL types (Mangan and Alon, 2003; Nitzan et al., 2017). Moreover, FFLs can display different logic gates for their targets' dual regulation: 'AND' logic, where both arms are required for regulation of C, or 'OR' logic, where one arm is sufficient (Beisel and Storz, 2010). When C is instead a set of multiple genes, a FFL is known as a multi-output FFL.FFLs are a main building block of transcriptional networks in *E. coli* (Alon, 2007).

sRNAs expression is controlled by TFs and σ factors but TFs are also known to be controlled by sRNAs. Therefore, mixed FFLs consisting of TFs and sRNAs are common

and can have two configurations: an sRNA can occupy the top layer, regulating the target both directly and indirectly through a TF. Alternatively, a TF can occupy the top node, regulating the target both directly and indirectly via a sRNA.

FFLs and their combination into more complex networks are exemplified by the regulation of the balance between ompC and ompF expression during osmotic stress (Nitzan et al., 2017). The switch between ompC and ompF expression is controlled by two coupled FFLs with a common regulator, OmpR (TF) at the top (Fig. 1.13 A). During osmotic stress, the OmpR-MicF-ompF module downregulates ompF and prevents its transcriptional leakage (OmpR represses *ompF* and activates MicF that also represses the target) (Ramani et al., 1994), whereas the second FFL, OmpR-MicC-*ompC* leads to delayed OmpC synthesis (MicC represses ompC and OmpR represses MicC while activating *ompC*) (Chen et al., 2004).

Feedback loops are closed loops in which a regulator controls the expression of its

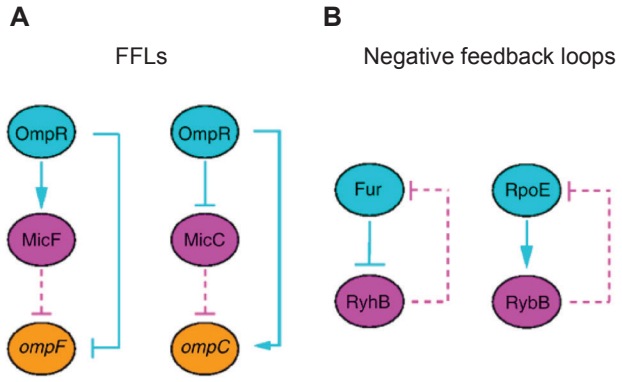

Fig. 1.13: Mixed regulatory circuits involving transcriptional regulation and post-transcriptional regulation by sRNAs. (A) Coherent feed-forward loops (FFLs). In response to osmotic stress, the OmpR transcription factor activates production of MicF sRNA, which represses OmpF protein synthesis (indirect arm). OmpR also directly represses *ompF* transcription (direct arm). At the same time, OmpR inhibits transcription of MicC sRNA, which inhibits translation of ompC. OmpR also activates directly the *ompC* transcription. The coupling of these two FFLs regulates the balance in OmpC and OmpF expression. (B) Negative feedback-loops: Fur represses transcription of RyhB sRNA, which in turn inhibits Fur synthesis. σ^E sigma factor (RpoE) activates RybB sRNA transcription, which in turn negatively regulates RpoE synthesis. Image adapted from Shimoni et al., 2007.

own gene. Positive feedback loops (the loop enhances sensitivity to the input signal) can amplify the signal and slow down the regulatory response (Maeda and Sano, 2006), whereas negative feedback loops can speed up a regulatory response and alter the relationship between signal and regulated genes (Beisel and Storz, 2010). All examples of feedback loops involving sRNAs are negative feedback loops, where a TF regulates a sRNA, which in turn regulates the TF. sRNAs must employ mixed FFLs to to feedback on their own transcription.

A positive-negative configuration, in which one regulation is positive whilst the other is negative, can induce a non-steady-state behaviour characterized by oscillatory expression dynamics. (Liu et al., 2011). A simple example of positive-negative configuration (Fig. 1.13 B) is the module σ^E-RybB, in which σ^E activates transcription of the sRNA gene *rybB* and RybB in turn represses σ^E synthesis via a homeostatic loop (Thompson et al., 2007).

A negative-negative configuration in which the TF and sRNA repress each other (e.g. Fur-RyhB) can induce two alternative internal states in response to different stimuli (bistabilty) (Nitzan et al., 2017; Liu et al., 2011).

These scenarios can greatly affect gene expression kinetics, and with bistability and oscillations, the post-transcriptional repression is more robust than by transcriptional repression alone (Liu et al., 2011).

1.10. Advantages of sRNA-mediated regulation

Although here I presented only a few selected examples, these types of regulatory network motifs are controlling every pathway and adaptive response in bacteria. The inclusion of sRNAs in gene expression control circuits is advantageous due to the inherent properties of sRNA-mediated regulation. With respect to the binary nature of the effect on gene expression (on or off), TFs and sRNAs may at a first glance seem redundant. However, based on the regulatory modes and the dynamics of responses

they endow, TFs and sRNAs are very distinct. The unique features of regulation by sRNAs, uncovered by quantitative studies, allow it to complement transcriptional regulation, and may explain the widespread physiological role of sRNAs in stress responses (Levine and Hwa, 2008).

Some conventional advantages of sRNA-mediated regulation include additional layers of control, reduced resource requirements, faster responses and unique regulatory features (Beisel and Storz, 2010). Quantitative effects of TFs and sRNA on their targets have been best described for negative regulation. Although TFs have a more dramatic effect on changes in gene expression than sRNAs (Hussein and Lim, 2012), regulation by sRNAs is faster than by TFs (Shimoni et al., 2007). It has been rationalised that fast-regulation by sRNAs stems from the fact that sRNAs act directly on the existing pool of transcripts with a decreased time-delay, and from the fact that compared to TFs, sRNA expression bypasses the need for translation (Shimoni et al., 2007). Conversely, TFs only alter the synthesis of future mRNAs. Thus, although a TF represses transcription of a target, each previous transcription event is still amplified by translation, because existing target mRNAs keep being translated (until they naturally decay) (Obzudak et al. 2002). sRNAs, on the other hand, can directly destabilize the transcript or inhibit translation, producing a faster (as well as additive) effect on expression of the target. It was also proposed that the dynamics of target expression are determined to a significantly greater extent by the steady states, clearance rates, and response curves of the regulators (Hussein and Lim, 2012). Either way, the consensus is that changes in sRNA expression lead to changes in target gene expression with a minimal delay (Nitzan et al., 2017), and indeed, compared to TFs, sRNAs have higher clearance rates (Hussein and Lim, 2012). This renders sRNAs particularly important in regulation of stress responses, which require rapid alterations in gene expression.

Although TF-mediated regulation is primarily determined by their binding to cognate DNA sequences, the major determinant in sRNA-mediated regulation is the expression level of the sRNA. Basically, the more target RNAs for a given sRNA are expressed,

the more sRNA molecules are expected to be in a duplex, hence unavailable to other targets, and vice-versa. Consequently, the sizes of free sRNA and free mRNA pools for a given sRNA-target pair are inversely correlated (Nitzan et al., 2017). This renders a role for weak sRNA targets in buffering the expression noise of a few preferential targets of a particular sRNA (Jost et al., 2013).

Moreover, sRNAs have been shown to establish gene expression thresholds for their targets, that are dependent on the transcription rate of the sRNA, reflective of changes in the external conditions. Given a sRNA with a unique target, if the transcription rate of the target RNA is lower than that of sRNA repressor, then most of the targets will pair with the sRNA, resulting in effective sRNA-mediated repression. Conversely, if the target is transcribed at higher rates than the sRNA, most sRNAs will base-pair with the target, but excess mRNAs can still be translated into protein. Consequently, the level of expressed protein will be linearly proportional to the difference between the two transcription rates, a behaviour known as the threshold-linear response (Levine and Hwa, 2008) (Fig. 1.14B). Below this dynamically defined threshold, sRNAs suppress random fluctuations and buffer translational noise and transient signals. Above the threshold, the effect of sRNA on target expression is negligible (Levine and Hwa, 2008). When expressed simultaneously, different target RNAs of the same sRNA regulon will compete for association with the sRNA. The outcome will depend on the relative levels of the targets and their affinities for interaction (Nitzan et al., 2017; Levine and Hwa, 2008). This scenario is complicated by the fact that sRNA-target RNA binding not only lead to target down-regulation, but also can lead to coupled degradation, thus depleting the regulator (Levine et al., 2007). This is illustrated by the activity of RNA sponges (Section 1.6). The competition of different mRNAs for an sRNA can be viewed as a sRNA-mediated crosstalk between mRNAs, a target-centric view (Bossi and Figueroa-Bossi, 2016), similar to the competing endogenous RNA hypothesis for miRNA regulation in eukaryotes (Salmena et al., 2011). Moreover, competition for Hfq (Section 1. 5) and other RBPs also has to be taken into account when estimating mutual effects of sRNAs in regulatory networks (Bossi and Figueroa-

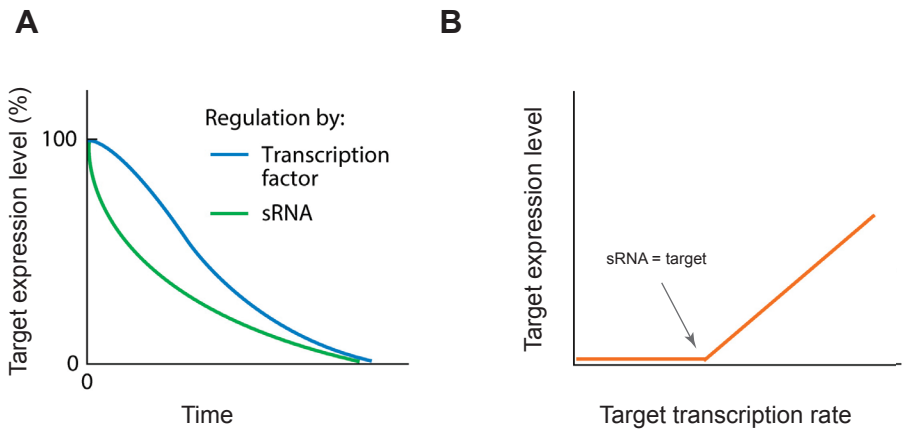

Fig. 1.14: Kinetic and quantitative considerations of sRNA mediated regulation. (A) Regulation by sRNAs is effective faster than that by TFs. Repression of a single gene expressed as level of target expression (y-axis) versus time (x-axis) (B) The threshold-linear response. The expression of target gene is effectively repressed as long as its transcription rate is lower than that of regulator sRNA. Below the threshold (indicated by arrow) at which sRNA transcription rate is equal to that of the target, sRNAs supress fluctuations and transient signals. Above the threshold, the effect of sRNA is minimal. Figure reproduced from Nitzan et al., 2017

Bossi, 2016).

Finally, the combination of TF- and sRNA-mediated regulation endows regulatory circuits with distinct dynamic properties and help reduce leakiness (Beisel and Storz, 2010). There are still factors that are absent in our current interpretation of sRNA regulatory networks. For instance, the recent enrichment with sRNA-mRNA interactions data is not sufficient to reveal the directionality and functionality of regulatory circuits, in absence of information on the outcome of pairing. Another challenge is the inclusion of effects rendered by RBPs and their dynamics of binding to RNA on the topology of sRNA networks.

2. Materials and methods

2.1. Bacterial strains and culture conditions

An overview of the bacterial strains used in this study is provided in Table 2.1.

The *E. coli* MG1655 or TOP10 strains served as parental strains. The *E. coli* K12 strain used for CLASH experiments, MG1655 *hfq*::HTF, was kindly provided by Jai Tree (Peter Doherty Institute, Melbourne). Hfq was chromosomally tagged in an MG1655 background with the dual-affinity tag HTF (containing His6, a TEV protease cleavage site, and a 3xFLAG) by allelic exchange using the pTOF series of vectors (Tree et al., 2014; Merlin et al, 2002). This strain was tested for normal Hfq-mediated regulation (Tree et al, 2014).

Table 2.1: *E. coli* strains used in study

Strain	Genotype	Reference
E.coli K12 MG1655	F- lambda- *rph*-1	Blattner et al., 1997
E.coli K12 MG1655 hfq::HTF	MG1655 *hfq*::HTF	Tree et al., 2014
N3431	Hfr(PO1) *lacZ*43(Fs) λ- *rne*-3071(ts) *relA*1 *spoT*1 *thiE*1	Babitzke et al, 1991
DH5α	*fhuA*2 *lac*(del) U169 *phoA gln*V44 80'Φ*lacZ*(del)M15 *gyrA*96 *recA*1 *relA*1 *endA*1 *thi*-1*hsdR*17	Taylor et al., 1993
TOP10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ *lacX*74 *recA*1 *araD*139 Δ(araleu)7697 *galU galK rpsL* (*StrR*) *endA*1 *nupG*	Invitrogen

Cells were grown in Luria-Bertani medium (LB) at 37°C under aerobic conditions with agitation at 200 rpm. The media were supplemented with antibiotics where required (Table 2.1) at the following concentrations: ampicillin - 100 µg/ml, chloramphenicol - 25 µg/ml, kanamycin - 50 µg/ml.

2.2. Construction of mRNA-superfolder GFP fusions

Table 2.2 lists all the plasmids, Table 2.3 all gene fragments, Tables 2.4 and 2.5 all primers used for cloning procedures in this work, respectively. To construct constitutively expressed, in-frame mRNA-*sfGFP* fusions for the fluorescence reporter studies, the 5'-UTR, start codon and first ~5 codons of target genes were cloned under the control of PLtetO-1 promoter in a pXG10-SF backbone as previously described (Urban and Vogel, 2007; Urban and Vogel, 2009; Corcoran et al, 2012). In cases where a membrane localisation signal was present, the respective sequence was excluded from the fusion design (e.g, OmpC). Derivatives of the target–GFP fusion plasmids harbouring seed mutations (SM) were generated using mutated gene-fragments (IDT) (Table 2.3).

For backbone preparation, 4 µg of pXG10-SF (4.7 kb) was digested with 0.9U NheI and 0.7U NsiI restriction enzymes (NEB) at 37ºC for 3 hours to remove the *lacZ i*nsert, and subsequently dephosphorylated with 1U rSAP (NEB) for 1 hour. After resolving the samples on a 1% agarose gel, the 4.1 kb band was gel-purified using the MinElute gel-extraction kit (Qiagen) according to manufacturer's instructions.

Table 2.2: . Plasmids used in study

Plasmid	Description	Comment	Origin, marker	Reference
pXG-0	PLtetO-1 luciferase expression vector	Cell autoflu-orescence control for GFP reporter studies	pSC101*, CmR	Urban and Vogel, 2007
pXG-1	PLtetO-1 GFP expression vector	Full-length GFP control for GFP reporter studies	pSC101*, CmR	Urban and Vogel, 2007
pxG10-SF	PLtetO-1 su-perfolder GFP fusion vector; lacZ insert	backbone for sfGFP plasmids	pSC101*, CmR	Corcoran et al, 2012
pXG10SF:: *ompC*	sfGFP report-er plasmid containing the ompC 5'UTR plus 12 amino acids	Constitutive* expression of OmpC-sfGFP	pSC101*, CmR	This study

pXG10SF:: ompCSM	sfGFP reporter plasmid containing	Constitutive* expression of OmpCSM-sfGFP	pSC101*, CmR	This study
pXG10SF:: *ompA*	sfGFP reporter plasmid containing	Constitutive* expression of OmpA-sfGFP	pSC101*, CmR	This study
pXG10SF:: ompASM	sfGFP reporter plasmid containing	Constitutive* expression of OmpASM-sfGFP	pSC101*, CmR	This study
pXG10SF::*tsx*	sfGFP reporter plasmid containing	Constitutive* expression of Tsx-sfGFP	pSC101*, CmR	This study
pXG10SF::tsx-SM	sfGFP reporter plasmid containing	Constitutive* expression of TsxSM-sfGFP	pSC101*, CmR	This study
pXG10SF:: *rpoE*	sfGFP reporter plasmid containing the full-length rpoE gene	Constitutive* expression of RpoE-sfGFP	pSC101*, CmR	This study
pBAD+1	pBADmycHisA expression vector with MCS removed from +1 to terminator	Control for overexpression studies	pBR322, AmpR	Tree et al, 2014
pBAD+1::MicC	MicC cloned at the +1 site of pBAD+1	Inducible expression of MicC	pBR322, AmpR	This study
pBAD+1::RybB	RybB cloned at the +1 site of pBAD+1	Inducible expression of RybB	pBR322, AmpR	This study
pBAD+1::MdoR	MdoR cloned at the +1 site of pBAD+1	Inducible expression of MdoR	pBR322, AmpR	This study
pBAD+1:: MdoRSM	MdoRSM cloned at the +1 site of pBAD+1	Inducible expression of MdoRSM	pBR322, AmpR	This study
pJET1.2/blunt	cloning vector with MCS for blunt-end PCR products	-	pBR322, AmpR	Thermo Fisher
pJET1.2::*malG*	pJET1.2 with full-length malG cloned in the MCS	Plasmid for MdoR primer extension ladder	pBR322, AmpR	This study

Table 2.3: Synthetic gene fragments cloned into pXG10-SF-derived backbone to generate mRNA-*sfGFP* fusions

Gene fragment	Sequence (5' → 3')
OmpCSM	gtttttATGCATTGCCGACTGATTAATGAGGGTTAATCACATACGTCAC-CGTATTTTAAGCAAATAAAGGCATATAACAGAGGGTTAATAACat-gAAAGTTAAAGTACTGTCCCTCCTGGTCCCAGCTAGCaaaaac
OmpA	tgacATGCATACATCGCCAGGGGTGCTCGGCATAAGCCGAAGA-TATCGGTAGAGTTAATATTGAGCAGATCCCCCGGTGAAGGAT-TTAACCGTGTTATCTCGTTGGAGATATTCATGGCGTATTTTGGATGA-TAACGAGGCGCAAAAAATGAAAAAGACAGCTATCGCGATTGCAGT-GGCACTGGCTGGTTTCGCTGCTAGCgtca
OmpASM	tgacATGCATACATCGCCAGGGGTGCTCGGCATAAGCCGAAGA-TATCGGTAGAGTTAATATTGAGCAGATCCCCCGGTGAAGGAT-TTAACCGTGTTATCTCGTTGGAGATATTCATGGCGTATTTTGGAT-GATAACGAGGCGCAAAAAATGAAAAAGACAGCTATCGCGATACGT-CACCGTGACGGCTGGTTTCGCTGCTAGCgtca
Tsx	tgacATGCATATCACCTGGATTGATAGTAAAAGTTTGCAACAAG-GGCGAAAGTCAGTACAATCCCCGCCCGAATGTGTGTAAACGT-GAACGCAATCGATTACGTAAATGATAGAACTGTGAAACGAAACATAT-TTTTGTGAGCAATGATTTTTATAATAGGCTCCTCTGTATACGAAATAT-TTAGAAACGCAATTTGCGCCTTTTTCACTCCCGCAAGGGAT-TTTCAAACAGTGGCATACATATGAAAAAAACATTACTGGCAGCTAG-Cgtca
TsxSM	tgacATGCATATCACCTGGATTGATAGTAAAAGTTTGCAACAAG-GGCGAAAGTCAGTACAATCCCCGCCCGAATGTGTGTAAACGT-GAACGCAATCGATTACGTAAATGATAGAACTGTGAAACGAAACATAT-TTTTGTGAGCAATGATTTTTATAATAGGCTCCTCTGTATACGAAATAT-TTAGAAACGCAATTTGCGCCTTTTTCACTCCCGCAAGGGAT-TTTCATTGTCACCGTATGATATGAAAAAAACATTACTGGCAGCTAG-Cgtca
RpoE	tgacATGCATGCAGTTAAATGGGCATTTCTACACAGATAATGCGATGT-TCAGATTCTGTAGACTTATAATGATAGATAATGATCCGTCTACAGCAT-GACAAACAAAAACAGATGCGTTACGGAACTTTACAAAAACGAGA-CACTCTAACCCTTTGCTTGCTCAAATTGCAGCTAATGGAGTGGCGT-TTCGATAGCGCGTGGAAATTTGGTTTGGGGAGACTTTACCTCG-GATGAGCGAGCAGTTAACGGACCAGGTCCTGGTTGAACGGGTC-CAGAAGGGAGATCAGAAAGCCTTTAACTTACTGGTAGTGCGCTAT-CAGCATAAAGTGGCGAGTCTGGTTTCCCGCTATGTGCCGTCGGGT-GATGTTCCCGATGTGGTACAAGAAGCTTTTATTAAAGCCTATCGTG-CGCTGGATTCGTTCCGGGGAGATAGCGCTTTTTATACATGGCTG-TATCGGATTGCTGTAAATACAGCGAAAAATTACCTGGTTGCT-CAGGGGCGTCGTCCACCTTCCAGTGATGTGGATGCCATT-GAAGCTGAAAACTTCGAAAGTGGCGGCGCGTTGAAAGAAAT-TTCGAACCCTGAGAACTTAATGTTGTCAGAAGAACTGAGACAGA-TAGTTTTCCGAACTATTGAGTCCCTCCCGGAAGATTTACGCATGG-CAATAACCTTGCGGGAGCTGGATGGCCTGAGCTATGCTAGCgtca

Table 2.4: Oligonucleotides used for the cloning procedures

#	Oligonucleotide	Sequence (5' → 3')	Purpose
672	ompC_NsiI_JVO-0428	GTTTTTATGCATTGC-CGACTGATTAATGAGG	PCR primers to amplify ompC for ligation into pXG10SF = pOmpC::gfp
673	ompC_NheI_JVO-0429	GTTTTTGCTAGCTGGGAC-CAGGAGG	PCR primers to amplify ompC for ligation into pXG10SF = pOmpC::gfp
675	p-ZE-CAT	TGGGATATATCAACGGTG-GT	sequencing primer for verification of ligation of PCR product into pXG10SF
681	ompC_pcr_rev	GCCGTAGTTACGACCGTA	PCR primers to check insertion of ompC by colony PCR
682	ompC_pcr_fwd	CGCCATTCCGCAATAATCT-TA	PCR primers to check insertion of ompC by colony PCR
683	Ec_malG_fwd	GCTGCTGATTAAACCGCT-GAC	PCR primer for malG ladder for primer extension
684	Ec_malG_rev	CCGTGACTCAGAGCACGAA	PCR primer for malG ladder for primer extension
710	pBAD+1_XbaI_fwd	AAAAATCTAGATTTGCCT-GGCGGCAGTAGCG	reverse primer to generate empty pBAD+1
711	pBAD+1_5P_rev	5P-TATGGAGAAACAGTAGA-GAG	fwd primer to generate empty pBAD+1
712	MdoR_pBAD_short	AAAGATGTTGTTCTGC-CAATG	rev primer for inverse PCR - pBAD+1
713	seq_pxg10SF_insert_fwd	GAAAGTTGGAACCTCT-TACGTG	fwd sequencing primer to check insertion into pXG10-SF backbone
715	sequencing_pBAD+1_fwd	ATGCCATAGCATTTTTATCC	fwd sequencing rimer to check insertion into pBAD+1 backbone
716	sequencing_pBAD+1_rev	CGCTACTGCCGCCAGG-CAAA	rev sequencing rimer to check insertion into pBAD+1 backbone

To prepare the insert, the target region of mRNA of interest was either amplified by PCR from *E. coli* genomic DNA (for OmpC) using primers with NsiI and NheI restriction sites (Table 2.4), or synthesised as gene fragments flanked by NsiI and NheI recognition sites at 5'- and 3' end, respectively (IDT; Table 2.3). 10-20 ng of PCR product or gene fragment was digested with NheI and NsiI in a 16 µl reaction for 1 hour at 37ºC and purified with a DNA Clean and Concentrator kit (Zymo Research)

according to kit instructions, then eluted in 8 µl DEPC water. For each mRNA-sfGFP fusion a small-scale ligation reaction (5 µl) using 12 ng of digested backbone and insert in a 3:1 insert:backbone molar ratio and 100 U T4 DNA Ligase (NEB) was incubated either for 3h at room temperature or 16°C overnight. The ligation reaction was transformed in TOP10 competent cells (Invitrogen, Table 2.1) by the heat-shock method and transformants were selected on chloramphenicol. Transformants were screened by restriction digest analysis and verified by Sanger sequencing (Edinburgh Genomics).

2.3. Construction of sRNA expression plasmids

For plasmid-borne over-expression of constructs, the sRNA gene of interest was cloned at the transcriptional +1 site under P*ara* control by amplifying the pBAD+1 plasmid (Table 2.4) by inverse PCR. The pBAD+1 template is derived from pBADmycHis A (Tree et al., 2014). The sRNA genes were synthesized as ultramers (IDT), which served as the forward primers (Table 2.5). To introduce mutations in the seed sequence of sRNAs (SM), the synthetic ultramer contained the desired mutation, namely the complement of the seed sequence. The reverse primer (oligo 712; Table 2.4) bears a monophosphorylated 5'-end to allow blunt-end ligation.

Table 2.5: Ultramers used for cloning sRNAs in pBAD+1 backbone; to be used with reverse oligo 712

Ultramer	Sequence (5' → 3')
pBAD+1+MdoR	AAAGATGTTGTTCTGCCAATGTTATGCCGCTGCACCCTCAACT-TACGTTATCCCAACTTGTGACTGTTATTCGGCGCTCCACGGAG-CGCCTTTTTTTTTTGCCTGGCGGCAGTAGCG
pBAD+1+MdoR SM	AAAGATGTTGTTCTGCCAATGAATACGGCGACGTCCCTCAACT-TACGTTATCCCAACTTGTGACTGTTATTCGGCGCTCCACGGAG-CGCCTTTTTTTTTTGCCTGGCGGCAGTAGCG
pBAD+1+RybB	GCCACTGCTTTTCTTTGATGTCCCCATTTTGTGGAGCCCAT-CAACCCCGCCATTTCGGTTCAAGGTTGATGGGTTTTTTGT-TTTGCCTGGCGGCAGTAGCG
pBAD+1+MicC	GTTATATGCCTTTATTGTCACAGATTTTATTTTCTGTTG-GGCCATTGCATTGCCACTGATTTTCCAACATATAAAAAGA-CAAGCCCGAACAGTCGTCCGGGCTTTTTTTTTTGCCTGGCGG-CAGTAGCG

The PCR cycling conditions were as recommended the polymerase manufacturer, with the exception that the annealing time was 3 min, and the annealing temperature was 50°C (Q5 DNA polymerase, NEB; with High GC buffer where necessary).The PCR reaction was digested with 10U DpnI (NEB) for 1h at 37ºC, then 5% of the PCR reaction was visualised by gel-electrophoresis for detection of the product. The rest of the PCR reaction was purified by ethanol precipitation.

The sRNA-pBAD linear PCR product was circularized by self-ligation. The blunt-end ligation was performed as above, using 20 ng PCR product and 300U T4 DNA ligase. Ligations were transformed in DH5α competent cells for propagation. Positive transformants (selected on ampicillin) were screened by sequencing. The control plasmid pBAD+1 was constructed similarly by self-ligation of the PCR product generated from oligonucleotides 710 and 711 (Table 2.4).

2.4. E. coli electroporation

The pBAD::sRNA and sfGFP plasmids constructed as in Section 2.2 (Plasmids pBAD::MdoR, pBAD+1, pBAD::MdoRSM, pBAD::rybB ;Table 2.2) were transformed in *E.coli* by electroporation. Cells were cultured in LB at 37ºC to OD_{600} of 2.0 and diluted 1:100 in fresh LB, and further grown to OD_{600} 0.3-0.4. Cultures (5 ml per electroporation) were centrifuged at 4ºC for 5 min at 3.5 krpm, then resuspended in 10-15 ml ice-cold 10% glycerol and centrifuged 10 min at 3.5 krpm. This step was repeated for a total of three washes. After the final wash, residual glycerol was removed and the cell pellet was resuspended in 10% glycerol to concentrate the cells 100-fold. 75 µl of this cell suspension were added to ice-cold, 0.2 cm cuvettes (Biorad) containing 10-20 ng plasmid and pulsed in a BioRad cell-pulser with the following settings: 200 ohm, 25 uF, 2.5 kV. Immediately after electroporation the cells were added to 1 ml warm LB and recovered for one hour at 37ºC with shaking at 300 rpm. Transformants were selected on antibiotics.

2.5. Radiolabeling of oligonucleotide probes

DNA oligonucleotides (Table 2.6) were labeled with 32P gamma-ATP with 10U T4 PNK (NEB) in 1XPNK buffer and 10 mM DTT at 37 °C. The reaction was purified using mini Quick Spin columns (Roche) to remove un-incorporated label as per manufacturer's instructions, yielding a final concentration of 1 µM 32P-end-labeled probe.

2.6. Northern Blot analysis

Cell pellets were resuspended in 100 µl GTC-Phenol and lysed by vortexing the cells for 5 min with 100 µl of acid-washed glass beads (425-600 µM diameter; Sigma). Total RNA was extracted from lysates by GTC-Phenol extraction (Tollervey et al, 1987).

Table 2.6: Oligonucleotides used for RNA probing

#	Oligonucleotide	Sequence (5' → 3')	Purpose
641	anti-malG14	GCGGCATAACATTGGCAGAAC	Northern probe for MdoR
643	JVO-0322 5S	CTACGGCGTTTCACTTCTGAGT-TC	Northern probe for 5S rRNA
644	anti-INT_0_2492	GGCAACGAAGGGTTCTACT	Northern probe for intergenic sRNA candidate
645	anti-INT_0_1913	GAAGACAGGCTTCGGGGTCA	Northern probe for intergenic sRNA candidate
646	anti-INT_0_1420	GGTGATGGCGTTCATAGCA	Northern probe for intergenic sRNA candidate
647	anti-INT_0_3041	GCCAGTCAGTTGCTTAACTGA	Northern probe for intergenic sRNA candidate
667	ompC3	GCCACTGCATACTGATAACCCTC	Northern probe for ompC; anneals to *ompC* 5'-UTR
670	micC_NB-JVO-1369	CAGTGGCAATGCAATGGCCCAA-CAG	Northern probe for MicC
678	malG_pe	GTGGAGCGCCGAATAACA	Oligo PE for 5'end mapping of MdoR / *malG*

703	revSF	GAATTGGGACAACTCCAGTG	Check GFP-fused constructs by sequencing and PCR (pxg10SF)
728	NB probe RybB_ JVO-1205	GTTGATGGGCTCCACAA	Northern probe for RybB
791	anti-malG14_ MdoR_seed_mut	CGCCGTATTCATTGGCAGAAC	Northern probe for MdoR SM
792	anti-RaiZ	CACGACGTGCTTCGCC	Northern probe for RaiZ homologue

For large RNA fragments, 10 µg of total RNA was resolved on a 1.25% BPTE- gel (pH 7) and transferred to a nylon membrane (HyBond N+, GEHealthcare) by capillarity. For short RNA fragments, 10 µg total RNA was separated on a 8% polyacrylamide TBE-Urea gel and transferred to a nylon membrane by electroblotting for four hours at 50 V. All membranes were UV-crosslinked in a Stratalinker at intensity 1200. Membranes were pre-hybridised in 10 ml of UltraHyb Oligo Hyb (Ambion) for one hour and probed with ^{32}P end-labeled DNA oligo for 12-18 hours in a hybridization oven at 42ºC. The sequences of the probes used for Northern blot detection are detailed in Table 2.6. Membranes were washed two times in 2xSSC with 0.5% SDS solution for 10 minutes, and visualized with a Phosphor imaging screen and FujiFilm FLA-5100 Scanner (IP-S mode). For detection of very abundant RNA species (5S rRNA) an autoradiography film GE Healtchcare) was used for exposure.

2.7. Primer extension analysis

One microgram total RNA was reverse-transcribed using SuperScript III reverse transcriptase (Invitrogen) using ^{32}P-radiolabelled oligonucleotides as primers (Table 2.6). Primers were added to the RNA and annealing was performed by heating the sample at 85ºC for three minutes and then snap chilling them on ice. The reverse-transcription (RT) was performed for one hour at 45ºC, followed by Exonuclease I and RNaseIf (NEB) (0.5 µl each) treatment for 30 minutes at 37ºC. Reactions were stopped by mixing with an equal volume of 2XRNA loading dye (NEB), 2 minutes incubation at 95 ºC and snap chill. The sequencing ladders were prepared with Sequenase v2.0 (USB/Affymetrix) according to specified instructions using as a template a plasmid

with the sequence that needs to be mapped. Samples were resolved on 6% PAA/8M TBE-urea gels and visualized using the FLA5100 phosphoimager system.

2.8. Pulse over-expression studies

Overnight MG1655 cultures containing pBAD::sRNA and empty pBAD+1 control plasmids were re-inoculated in fresh LB-ampicillin medium at a starting OD_{600} of 0.05, and grown aerobically at 37ºC to OD_{600} 0.4. The cultures were then rapidly transferred to a shaking water bath pre-warmed to 37ºC. Cultures were supplemented with 0.2%(w/v) L-arabinose to induce transcription from the pBAD promoter, and cells were rapidly collected by filtration and flash-frozen in liquid nitrogen at different time-points after induction and stored at -80ºC. RNA was extracted from three biological replicate time-series, followed by RNASeq library preparation and next generation sequencing.(Section 2.14) and analysis of differentially expressed genes (DeSeq2 - Section 2.15).

For pBAD::RybB and pBAD::MdoRSM, two independent L-arabinose induction experiments (each accompanied by repeats of pBAD+1 control induction) were followed by RT-qPCR quantification of differentially expressed genes.

2.9. Western Blot analysis

Two OD_{600} units of *E.coli* cells were pelleted by centrifugation and resuspended in 50 µl TN150 buffer and lysed for 10 min at 95 ºC, then incubated for 5 minutes at 37ºC. Lysates were treated with 2 U Turbo DNase (Ambion) for 30 minutes at 37ºC. The protein content of the lysates was quantified by nanodrop (NanoDrop1000). A lysate volume containing 80 µg of protein was denatured for 10 minutes at 65ºC in 20 µl 1X LDS loading buffer (NuPAGE) and resolved on 4-12% gradient polyacrylamide gels (NuPAGE) alongside a broad-range protein standard (NEB, P7712). The proteins were transferred to nitrocellulose membranes using the iBlot (Invitrogen) and program 8. The membranes were blocked for one hour in blocking solution (5% non-

fat milk in PBST (1X phosphate saline buffer, 0.1% Tween-20). Antibody probing was performed in blocking solution as described below. Where necessary, after two five-minute washes in PBST, membranes were blotted with secondary antibodies. Finally, after three 10-minute PBST washes, the membranes were rinsed once with 1xPBS and proteins were visualised using Pearce enhanced chemiluminescence (ECL, ThermoFisher) according to manufacturer's instructions.

Antibody	Usage	
anti-ompC (Biorbyt; Rabbit)	1:1000	overnight, 4 ºC
anti-GFP (Roche; Mouse)	1:5000	1h, Room Temperature
anti-GroEL(Abcam; Rabbit)	1:300,000	2h, Room Temperature
Goat anti-Rabbit	1:5,000	1h, Room Temperature
Rabbit anti-Mouse	1:10,000	1h, Room Temperature

2.10. Transient inactivation of RNase E

The *E.coli rne*-3071 TS (Table 2.1) and parental MG1655 strains were grown in LB medium at 30ºC to an OD_{600} of 1.5, then shifted to 43ºC for 30 min to inactivate RNase E. Cells were harvested by rapid filtration (0.45 µM filters, Merck Millipore) and flash-frozen in liquid nitrogen.

2.11. RT-qPCR

Total RNA was extracted with the guanidium thiocyanate method as previously described (Tollervey et al., 1987). 12.5 ug RNA was treated with 2U Turbo DNase (Ambion) for 1 hour at 37ºC in a 10 µl reaction in the presence of 2U SuperaseIn RNase inhibitor (Ambion). The DNase was inactivated by 10 minutes incubation at 75ºC. Reverse transcription (RT) was performed in a single reaction for all target genes of interest using a mix of gene-specific RT primers (Table 2.7) at 3.5 µM concentration each. After addition of 2.5 µl RT primers mix, the RNA and primers were denatured at 70ºC for 3 min, then snap chilled and incubated on ice for 5 min.

RT was performed for 1h at 55ºC with SuperScript III (Invitrogen) using 5 µl of RNA-RT primers mix in 10 µl final volume (100U Superscript III, 2.5 mM DTT, 1xFS Buffer, 0.75 mM dNTPs) in the presence of 1U RNasin (Promega). A no reverse transcriptase (-RT) control, in which Superscript III was replaced with water was performed for all RTs. RT was followed by treatment with 5U RNase H for 30 min at 37ºC to remove the RNA from the RNA-cDNA duplexes. The cDNA was diluted 10-fold with DEPC water and stored at -20ºC.

Table 2.7: Oligonucleotides used for RT-qPCR analysis*

#	Oligonucleotide	Sequence (5' → 3')
779	gfpSF RT 2	GCGTCTTGTAGGTCCCGTCAT
780	recA RT	GCCGTAGAAGTTGATACCTTCGCC
781	rybB RT	ACAAAAAACCCATCAACCTTGAACCG
782	GFPsf_F	TTCCATGGCCAACACTTGTCA
783	GFPsf_R	TAACCTTCGGGCATGGCACT
784	recA 1 F_qPCR	GTATCGGCGCGGTGAAAGAG
785	recA 1 R_qPCR	TTAAACGGCGCAGCGATTTT
786	MdoR 1 FpPCR	TTATGCCGCTGCACCCTCAA
787	MdoR 1 RpPCR	TGGAGCGCCGAATAACAGTCA
788	MdoR SM F qPCR	AATACGGCGACGTCCCTCAA
789	rybB 1 F qPCR	TTTGATGTCCCCATTTTGTGGAG
790	rybB 1 R qPCR	ACCCATCAACCTTGAACCGAAA
806	ompC RT	GTCAGTGTTACGGTAGGTCGCG
807	rseA RT	CGCCTGCTCAAACTGAAGCTG T
808	ryeA RT	GACCACCACGCTTTTTATTGACCA T
804	ryeA1F_qPCR	GGAGAGAGCCGTGCGCTAAA
805	ryeA1R_qPCR	CTGGAAAACCTGGCGTCGTC
810	rseA_R_qPCR	TCAGAGTGGAGTCGGCGTTG
811	ompA_R_qPCR	TGCGCCTCGTTATCATCCAA
815	ompA RT	CCAGGTGTTATCTTTCGGAGCGG
816	rpoE RT	GGAACGAATCCAGCGCACGATA
817	rpoE_R_qPCR	GACGGCACATAGCGGGAAAC
818	ompC _F_qPCR	TGGCATTCGCAGGTCTGAAA
819	ompC _F_qPCR	CCACCGAATTCTGGCAGTACG
820	rpoE_F_qPCR	CGGGTCCAGAAGGGAGATCA
821	ompA_F_qPCR	AGCAGATCCCCCGGTGAAG

Quantitative PCR was performed using the Brilliant III UltraFast SYBR Green QPCR Master Mix (Agilent): 3 µl of cDNA was mixed with 7 µl of qPCR master mix containing the PCR primers listed in Table 2.7 (final concentration 300 nM) according to manufacturer's instructions. The PCR was run on a LightCycler 480 (Roche) with the program parameters recommended by Agilent: pre-incubation: 95ºC-5s; 45 cycles amplification with single acquisition mode: 95ºC-5s, 60ºC-10s; and melting curve: 65ºC-60s, 95ºC(0.11 ramp rate with 5 acquisitions per ºC, continuous). The following analyses were performed with the IDEAS2.0 software, both with default settings: Absolute Quantification/Fit Points for Cp determination and Melt Curve Genotyping to verify the specificity of the PCR. To calculate the fold-change relative to the control, the 2^ddCp method was employed, using *recA* as the reference gene, because it was stably expressed in the tested conditions. Negative controls such as -RT or no template control were used throughout, and the qPCR for all samples was performed in technical triplicate for each plate. Experiments were performed for minimum two biological replicates, and the mean and standard error of the mean were computed. Samples with technical triplicate standard deviations of Cp larger than 0.3 were discarded from the analysis.

2.12. GFP reporters of sRNA-mediated regulation

A two-plasmid system was used to express each sRNA, and mRNA-*sfGFP* fusions (Section 2.3). The two types of plasmids were co-transformed in *E. coli* TOP10 cells by electroporation (Section 2.4) and cells were maintained on dual selection with ampicillin and chloramphenicol. In TOP10 cells, the mRNA-*sfGFP* constructs are constitutively expressed, whereas sRNA expression requires L-arabinose induction. The expression of sfGFP-fused targets in the presence or absence of sRNAs was quantified at the protein level, by plate reader experiments and at the RNA level, by RT-qPCR (Section 2.11).

For the plate reader experiments, a single colony of bacterial strain harbouring a sRNA-target-sfGFP combination was inoculated in a 96-well Flat Bottom Transparent

Polystyrene plate (Fisher Scientific) with lid (Thermo Scientific) and cultured overnight at 37ºC in 100 µl LB supplemented with antibiotics and 0.2% L-arabinose to induce expression of sRNAs from the P_{ara} promoter.

Next day, each overnight inoculum was diluted 1:100 by serial dilution, in triplicate, in LB with freshly prepared 0.2% L-arabinose to a final volume of 100 µl. Cultures were grown in a 96-well plate in an Infinite 200 Pro plate reader (Tecan) controlled by i-control software (Tecan) for 192 cycles at 37ºC with 1 min orbital shaking (4 mm amplitude) every 5th minute. To monitor optical density over time, the following parameters were used: wavelength 600 nm, bandwidth 9 nm. Fluorescence was monitored with excitation wavelength 480 nm, bandwidth 9 nm and emission wavelength 520 nm, bandwidth 20 nm. Measurements were recorded at 5 minute intervals, by top reading. Raw data was processed using custom python scripts that I wrote following guidance from previous reports (Urban and Vogel, 2007). First, the range of linearity of increase of fluorescence with OD_{600} was identified for all individual triplicates. Only the OD_{600} range common to all triplicates was considered for further analysis. For each set of triplicates, the mean fluorescence was calculated at each OD_{600}. To correct for background and cell autofluorescence, the fluorescence mean of a strain with plasmid pXG-0 was subtracted from all strains with GFP plasmids at the equivalent OD_{600} (Table 2.2). Ultimately, a curve was generated for each sample plotting the background-corrected fluorescence (GFP) versus OD_{600}. The experiments were performed for three biological replicates, and mean values and standard error of the means calculated for each strain.

To quantify the effect of sRNA expression on GFP reporters at the RNA level, the sRNAs were pulse-overexpressed as described above (Section 2.6) by inducing sRNA transcription for 20 minutes with 0.2% L-arabinose. RT-qPCR (Section 2.11) was performed to detect the GFP region of the mRNA-sfGFP fusion and pXG-1.

2.13. Preparation of CLASH libraries

E. coli expressing the chromosomal Hfq-HTF were grown overnight in LB at 37°C

with shaking (200 rpm.), diluted to starter OD_{600} 0.05 in fresh LB, and re-grown with shaking at 37°C in 750 ml LB. A volume of culture equivalent to 80 OD_{600} per ml was removed at the following time-points (OD_{600}): 0.4, 0.8, 1.2, 1.8, 2.4, 3.0 and 4.0, respectively, and immediately subjected to UV (254 nm) irradiation for 22s in the Vari-X-linker (Van Nues et al, 2017). The cells were harvested by rapid filtration with 0.45 µM nitrocellulose filters (Merck Millipore) and flash-frozen in liquid nitrogen. Cells were resuspended from filters by three washes with ~15 ml ice-cold phosphate-buffered saline (PBS), and centrifuged for 10 min at 4000 rpm at 4 C.

During the optimization of CLASH for *E. coli,* Sander Granneman and myself generated five CLASH libraries. The last two showed very high read complexity and were used for further analysis. Cell pellets were lysed by bead-beating in 1 volume per weight TN150 buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.1% NP-40, 5 mM β-mercaptoethanol) in the presence of protease inhibitors (Roche), and 3 volumes 0.1 mm Zirconia beads (Thistle Scientific; 0.1 mm diameter), by performing 5 cycles of 1 minute vortexing followed by 1 minute incubation on ice. One additional volume of TN150 buffer were added and chromosomal DNA was removed with RQ1 DNase I treatment (Promega) for 30 min on ice. Two-additional volumes of TN150 were added and mixed with the lysates by vortexing. The lysates were centrifuged for 20 minutes at 4000 rpm at 4ºC and subsequently clarified by a second centrifugation step at 13.4 krpm, for 20 min at 4ºC.

The Hfq-HTF-RNA complex purification and library preparation was essentially performed as described (Granneman et al, 2009). Cell lysates were incubated with 50 µl of pre-equilibrated M2 anti-FLAG beads (Sigma) for 2 hours at 4°C. The anti-FLAG beads were washed three times 10 min with 1 ml TN1000 (50 mM Tris pH 7.5, 0.1% NP-40, 5 mM β-mercaptoethanol, 1 M NaCl) and rinsed three times with TN150 without protease inhibitors (50 mM Tris pH 7.5, 0.1% NP-40, 5 mM β-mercaptoethanol, 150 mM NaCl). For TEV cleavage, the beads were resuspended in 250 µl of TN150 buffer without protease inhibitors, and incubated at room temperature for 2 hours, with rotation with 10 µl of home-made recombinant GST-TEV protease. The TEV

eluates were then incubated with a 1:50 dilution preparation of RNaceIt (RNase A and T1 mixture; Agilent) for exactly 5 minutes at 37ºC, after which they were mixed with 0.4g GuHCl (Sigma). The samples were then transferred to 100 µl Nickel-NTA agarose beads (Qiagen), equilibrated with wash buffer 1 (6 M GuHcl, 0.1% NP-40, 300 mM NaCl, 50 mM Tris pH 7.8, 10 mM Imidazole, 5 mM beta-mercaptoethanol). Binding was performed at 4ºC overnight with rotation. The next day, the beads were transferred to Pierce SnapCap spin columns (Thermo Fisher), washed 3 times with wash buffer 1 and 3 times with 1xPNK buffer (10 mM MgCl2, 50mM Tris pH 7.8, 0.1% NP-40, 5 mM beta-mercaptoethanol). The washes were followed by on-column TSAP incubation (Thermosensitive alkaline phosphatase, Promega) treatment for 1 h at 37ºC with 8 U of phosphatase in 60 µl of 1xPNK, in the presence of 80U RNasin (Promega). The beads were washed once with 500 µl wash buffer 1 and three times with 500 µl 1xPNK buffer. To add 3'-linkers (App-PE - Table 2.8), the Nickel-NTA beads were incubated in 80 µl 3'-linker ligation mix with (1 X PNK buffer, 1 mM 3'-adapter, 10% PEG8000, 30U Truncated T4 RNA ligase 2 K227Q (NEB), 60U RNasin). The samples were incubated for 4 hours at 25ºC. The 5'-ends of bound RNAs were radiolabeled with 30U T4 PNK (NEB) and 3 µl ^{32}P-γATP (Perkin Elmer) in 1xPNK buffer for 40 min at 37ºC, after which 100 mM cold ATP (Roche) was added to a final concentration of 1 mM, and the incubation prolonged for another 20 min to complete 5'-end phosphorylation. The resin was washed three times with 500 µl wash buffer 1 and three times with equal volume of 1xPNK buffer. For on-bead 5'-linker ligation, the beads were incubated 16h at 16ºC in 1xPNK buffer with 40U T4 RNA ligase I (NEB), and 1 µl 100 µM L5 adapter (Table 2.8), in the presence of 1mM ATP and 60U RNasin (Promega). The Nickel-NTA beads were washed three times with wash buffer 1 and three times with buffer 2 (50 mM Tris–HCl pH 7.8, 50 mM NaCl, 10 mM imidazole, 0.1% NP-40, 5 mM β-mercaptoethanol).

The protein-RNA complexes were eluted in two steps in new tubes with 200 µl of elution buffer (wash buffer 2 with 250 mM imidazole). The protein-RNA complexes were Tri-chloro acetic acid (TCA) precipitated on ice by adding TCA to a final concentration of 20%, followed by a 20-minute centrifugation at 4ºC at 13.4 krpm. Pellets were washed

with 800 µl acetone, and air dried for a few minutes in the hood. The protein pellet was resuspended and incubated at 65ºC in 20 µl 1x NuPage loading buffer (Novex), resolved on 4–12% NuPAGE gels, and visualised by autoradiography. The cross-linked proteins-RNA were cut from the gel incubated with 160 µg of Proteinase K (Roche) in 600 µl wash buffer 2 supplemented with 1% SDS and 5 mM EDTA at 55ºC for 2-3 hours with mixing. The RNA was subsequently extracted by phenol-chloroform extraction and ethanol precipitated. The RNA pellet was directly resuspended in RT buffer and was transcribed in a single reaction with the SuperScript III system (Invitrogen) according to manufacturer's instructions using the PE_reverse oligo as primer (Table 2.8). The cDNA was purified with the DNA Clean and Concentrator 5 kit (Zymo Research) and eluted in 11 µl DEPC water. Half of the cDNA (5 µl) was amplified by PCR using Pfu Polymerase (Promega) with the cycling conditions (95°C for 2 min; 20-24 cycles: 95°C for 20s, 52°C for 30s and 72°C for 1 min; final extension of 72°C for 5 min). The PCR primers are listed in Table 2.8. PCR products were treated with 40U Exonuclease 1 (NEB) for 1 h at 37ºC to remove free oligonucleotide and purified by ethanol precipitation/ or the DNA Clean and Concentrator 5 kit (Zymo Research). Libraries were resolved on a 2% MetaPhor agarose (Lonza) gel and 175-300bp fragments were gel-extracted with the MinElute kit (Qiagen) according to manufacturer's instructions. All libraries were quantified on a 2100 Bionalyzer using the High-Sensitivity DNA assay. Individual libraries were pooled based on concentration and barcode sequence identity. Paired-end sequencing (75 bp) was performed by Edinburgh Genomics on a Illumina HiSeq 4000 platform.

2.14. RNAseq library generation

E. coli MG1655 was cultured and harvested as described (Section 2. 13). Total RNA was extracted using the Guanidium thiocyanate phenol method (Tollervey et al, 1988). RNA integrity was assessed with the Prokaryote Total RNA Nano assay on a 2100 Bioanalyzer (Agilent). Contaminating genomic DNA was removed by incubating 10 µg of total RNA with 2U Turbo DNase (Ambion) in a 50 µl final volume for 30 minutes at 37ºC in the presence of 10 U SuperaseIn RNase Inhibitor (Ambion). RNA was

subsequently phenol-chloroform extracted and precipitated using 2.5V of 96% ethanol at -80ºC for 20 minutes. After washing the pellet with 70% ethanol, it was briefly air-dried and resuspended in 10 µl of DEPC-treated water. Ribosomal RNA was removed with the Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria) (Illumina) according to manufacturer's instructions. Successful rRNA depletion was verified on the 2100 Bioanalyzer as indicated by the disappearance of 16S and 23S rRNA species.

The RNA was fragmented by incubation for 5 min at 95ºC in the presence of Superscript III buffer (Invitrogen) followed by a five minute incubation on ice. Reverse-transcription (RT) was performed with Superscript III (Invitrogen) in 20 µl reactions according to manufacturer's procedures using 250 ng of ribosomal RNA depleted RNA and 2.5 µM random hexamers (oligo 73, Table 2.8). After the RT, the RNA and unannealed primers were degraded using 20U of Exonuclease I (NEB) and 50U RNaseIf (NEB) and the cDNA was purified with the DNA Clean & Concentrator 5 kit (Zymo Research) as per manufacturer's procedures. Ligation of the 5' adapter (oligo 39, Table 2.8) to the cDNA was performed using CircLigase II (EpiCentre) for 6 hours at 60ºC, followed by a 10 minute incubation at 80ºC to inactivate the enzyme. The cDNA was purified with the DNA Clean & Concentrator 5 kit (Zymo Research).

Half of the cDNA library was PCR amplified using Pfu polymerase (Promega) using the P5 forward PCR oligonucleotide and barcoded BC reverse oligonucleotides (200 nM; Table 2.8; 95°C for 2 min, 95°C for 20s, 52°C for 30s and 72°C for 1 min, and a final extension of 72°C for 5 min. 20 cycles of amplification).

PCR products were treated with Exonuclease 1 (NEB; 40U) for 1 h at 37ºC and purified by ethanol precipitation. Libraries were resolved on a 2% MetaPhor agarose (Lonza) gel 200-500 bp fragments were gel-extracted using the MinElute kit (Qiagen) according to manufacturer's instructions. All libraries were quantified on a 2100 Bionalyzer using the High-Sensitivity DNA assay. Individual libraries were pooled in equimolar amounts. Paired-end sequencing (75 bp) was performed by Edinburgh Genomics on a Illumina HiSeq 4000 platform.

Table 2.8: Oligonucleotides used for library preparation

CLASH L5 5’ adapters:	
BARCODE SET A	
L5Aa	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrUrArArGrCrN-OH-3'
L5Ab	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrArUrUrArGrCrN-OH-3'
L5Ac	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrGrCrGrCrArGrCrN-OH-3'
L5Ad	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrCrGrCrUrUrArGrCrN-OH-3'
BARCODE SET B	
L5Ba	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrArGrArGrCrN-OH-3'
L5Bb	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrGrUrGrArGrCrN-OH-3'
L5Bc	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrCrArCrUrArGrCrN-OH-3'
L5Bd	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrUrCrUrCrUrArGrCrN-OH-3'
BARCODE SET C	
L5Ca	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrCrUrArGrCrN-OH-3'
L5Cb	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrUrGrGrArGrCrN-OH-3'
L5Cc	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrArCrUrCrArGrCrN-OH-3'
L5Cd	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrGrArCrUrUrArGrCrN-OH-3'
BARCODE SET D	
L5Da	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrCrGrUrGrArUrN-OH-3'
L5Db	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrGrCrArCrUrArN-OH-3'
L5Dc	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrUrArGrUrGrCrN-OH-3'
L5Dd	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrArUrCrArCrGrN-OH-3'
BARCODE SET E	
L5Ea	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrCrArCrUrGrUrN-OH-3'
L5Eb	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrGrUrGrArCrArN-OH-3'
L5Ec	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrUrGrUrCrArCrN-OH-3'
L5Ed	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrN-rNrNrArCrArGrUrGrN-OH-3'
RNAseq 5’-adapter:	

P5_phospho_adapter (39):	5'P-AGATCGGAAGAGCGTCGTGTAGGG-3SpC3
CLASH 3' adapters:	
App_PE:	5'App-NAGATCGGAAGAGCACACGTCTG-ddC 3'
PCR oligonucleotides:	
P5 (Forward)	5'-AATGATACGGCGACCACCGAGATCTACACTCT-TTCCCTACACGACGCTCTTCCGATCT-3'
P3 (Reverse; BC1-BC6)	5'-CAAGCAGAAGACGGCATACGAGAT(BC 6nt barcode) GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
BC1	5'-CAAGCAGAAGACGGCATACGAGATCGTGATGT-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
BC2	5'-CAAGCAGAAGACGGCATACGAGATACATCGGTGACT-GGAGTTCAGACGTGTGCTCTTCCGATCT-3'
BC3	5'-CAAGCAGAAGACGGCATACGAGATGCCTAAGT-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
BC4	5'-CAAGCAGAAGACGGCATACGAGATTGGTCAGT-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
BC5	5'-CAAGCAGAAGACGGCATACGAGATCACTGTGT-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
BC6	5'-CAAGCAGAAGACGGCATACGAGATATTGGCGTGACT-GGAGTTCAGACGTGTGCTCTTCCGATCT-3'
RT oligonucleotides:	
PE_reverse (CLASH)	5'-CAGACGTGTGCTCTTCCGATCT-3'
PE_solexa_hexamer (RNA-seq; 73)	5'-CGTGTGCTCTTCCGATCTNNNNNN-3'

2.15. Bioinformatics analyses

Pre-processing of the raw sequencing data

Raw sequencing reads were processed using a pipeline developed by Sander Granneman, which uses tools from the pyCRAC package (Webb et al., 2014). The entire pipeline is available at https://bitbucket.org/sgrann/). The CRAC_pipeline_PE.py pipeline first demultiplexes the data using pyBarcodeFilter.py and the in-read barcode sequences found in the L5 5' adapters. Flexbar then trims the reads to remove 3'-adapter sequences and poor quality nucleotides (Phred score <23). Using the random nucleotide information present in the L5 5'-adaptor sequences, the reads were collapsed to remove potential PCR duplicates. The reads were then mapped to the *Escherichia coli* MG1655 genome with Novoalign (www.novocraft.com). To

determine which genes the reads overlapped with we generated an annotation file in the Gene Transfer Format (GTF). This file contains the start and end positions of each gene on the chromosome as well as what genomic features (i.e. sRNA, protein-coding, tRNA) it belongs to. To generate this file, we used the Rockhopper software (Tjaden, 2005) that took *E. coli* RNA-seq data generated by Christel Sirocchi, a minimal GTF file obtained from ENSEMBL (without UTR information). The resulting GTF file contained information not only on the coding sequences, but also complete 5' and 3' UTR coordinates. PyReadCounters then used the novoalign output file and the GTF file to count the total number of unique cDNAs that mapped to each gene.

Normalization steps

To normalise the read count data and to correct for differences in library depth between time-points, we calculated Transcripts Per Million reads (TPM) for each gene. Briefly, for each time-point the raw counts for each gene was first divided by the gene length and then divided by the sum of all the values for the genes in that time-point to normalize for differences in library depth. Subsequently, only genes with a minimum of 5 TPM in all datasets compared were used. The TPM values for each OD_{600} studied were divided by the TPM values of the first sample (OD_{600} 0.4). Thus, the fold-change starts at 1 for all samples in the OD_{600} series. These ratios were then $\log_2$-normalized. The $\log_2$-normalized fold-changes were used to compare RNAseq and Hfq-crosslinking profiles among samples, and to perform k-means clustering with the python sklearn. cluster.KMeans class.

Hfq-binding coverage plots

The pyBinCollector tool was used to generate Hfq cross-linking distribution plots over genomic features. First, PyCalculateFDRs.py (Webb et al, 2014) was used to identify the significantly enriched Hfq-binding peaks (min. 10 reads, minimum 20 nt intervals). Next, pyBinCollector was used to normalize gene lengths by dividing their sequences into 100 bins, and calculate nucleotide densities for each bin. To generate

the distribution profile for all genes individually, we normalized the total number of read clusters (assemblies of overlapping cDNA sequences) covering each nucleotide position by the total number of clusters that cover the gene,

Motif searches were performed with pyMotif.py using the significantly enriched Hfq-binding peaks. The 4-8 nucleotide k-mers with Z-scores above a set threshold were used for making the motif logo with the k-mer probability logo tool (Wu and Bartel., 2017) with the -ranked option (http://kplogo.wi.mit.edu/).

Analysis of hybrid reads

Chimeric reads were identified using the hyb package using default settings (Travis et al., 2014). To apply this single-end specific pipeline to our paired-end sequencing r data, we joined forward and reverse reads using FLASH, which merges overlapping reads into a single contig. These were then analysed using hyb. The -anti option for the hyb pipeline was used to be able to use a genomic *E. coli* hyb database, rather than a transcript database. Uniquely annotated hybrids (.ua.hyb) were used in subsequent analyses. To visualise the hybrids in the genome browser, the .ua.hyb output files were converted to the GTF format.

To generate distribution plots for the genes to which the chimeric reads mapped, the parts of the chimeras were clustered with pyClusterReads.py and bedtools (intersectBed) was used to remove clusters that map to multiple regions. To produce the coverage plots with pyBinCollector, each cluster was counted only once, and the number of reads belonging to each cluster was ignored.

Differential expression analyses

For the differential expression analyses DESeq2 was used (Love et al.,2014). Three MdoR pulse-overexpression datasets were compared to three pBAD Control overexpression datasets. Only differentially expressed genes were selected that had

an adjusted p-value of 0.05 or lower were considered significant.

Multiple sequence alignments and conservation analyses

The homologous sequences of MdoR in other entrobacteria were retrieved by BLAST. JalView was used for the multiple sequence alignments, using the MAFFT algorithm (Waterhouse et al., 2009). The distance tree was generated based on percentage identity of the input sequences. To predict MdoR targets in three bacterial species, CopraRNA was used at default parameters (Wright et al., 2014).

3. Global analyses of Hfq-associated RNA interactions

3.1. Introduction

In *E. coli*, hundreds of sRNAs have been discovered, but only a few have been characterized (Gottesman and Storz., 2011). The initial step in unravelling the roles of sRNAs is to capture their targets, for which many computational and *in vivo* methods have been developed (Section 1.8).

A main aim of this study was to uncover novel Hfq-mediated sRNA-target RNA interactions for the first time, at a global scale, using CLASH (Section 1.8; Helwak and Tollervey, 2014). However, during the course of this study, two similar approaches were adapted in *E. coli* - RIL-seq, that captures Hfq-associated RNA-RNA interactions (Melamed et al., 2016), and RNase E CLASH (Waters et al., 2017), that captures RNAse E - associated RNA-RNA interactions.

Nonetheless, the experimental design of the Hfq-CLASH experiment presented here is unique compared to the other studies, and aimed to cover a continuum of growth stages - to monitor the dynamics of Hfq binding to sRNAs and mRNAs, but also to increase the chances of recovery of sRNA-mRNA interactions representative for many conditions of growth.

Noteworthy, previous attempts to perform CLASH on Hfq have identified very few hybrids (~0.001% of mapped reads) (Waters et al., 2017). The frequency of hybrid reads recovery was much improved in RNase E CLASH. It was proposed that this was due to the transiency of Hfq association with sRNA-target RNA pairs – i.e sRNA-mRNA duplexes formation is coupled with the immediate dissociation of the RNAs from Hfq (Tree et al., 2014) and rapid transfer to RNase E (Waters et al., 2017).

Here I show that the CLASH protocol can be successfully applied on Hfq. The optimizations that were introduced during my PhD (i.e. shorter incubation times and controlling the RNAse digestion step more carefully) increased the overall recovery of chimeras over a 1000-fold.

The results presented in this Chapter showed that the Hfq binding to individual sRNAs and mRNAs is dynamic, growth-dependent. Moreover, analyses of the duplexes chaperoned by Hfq revealed complex sRNA networks and provided more examples of non-canonical mechanisms employed by sRNAs: sRNA-sRNA interactions and sRNA biogenesis from mRNAs. Many of the findings had little overlap with interactions uncovered by RIL-seq, thus they greatly expand the current knowledge of sRNA circuits.

3.2. Growth-dependent dynamics of Hfq-RNA interactions

To capture the dynamics of Hfq binding to sRNAs and mRNAs and to determine its correlation with changes in RNA levels during growth, and to identify the spectrum of sRNA-target interactions with roles during growth, CLASH (Section 1.8) was applied in *E. coli* cells harvested at different cell densities as measured by OD_{600} (Fig. 3.1 A). Seven time-points were analysed: OD_{600} 0.4, 0.8, 1.2, 1.8, 2.4, 3 and 4 (Fig. 3.1 B). These OD_{600} are representative of the main growth phases: exponential and stationary, but also include the transitions between them. The autoradiogram in Fig 3.1 C shows that Hfq cross-linking to RNA at each OD_{600} was comparable. The CLASH data was accompanied by RNAseq data collected under the same conditions, that provides information on the total RNA levels at each condition.

The sequencing of the RNAs bound by Hfq generated two types of reads. The single reads were used to generate Hfq-binding maps across the transcriptome. The chimeric reads, generated by ligation of RNA-ends of Hfq-bound duplexes, generally represent RNA-RNA interactions.

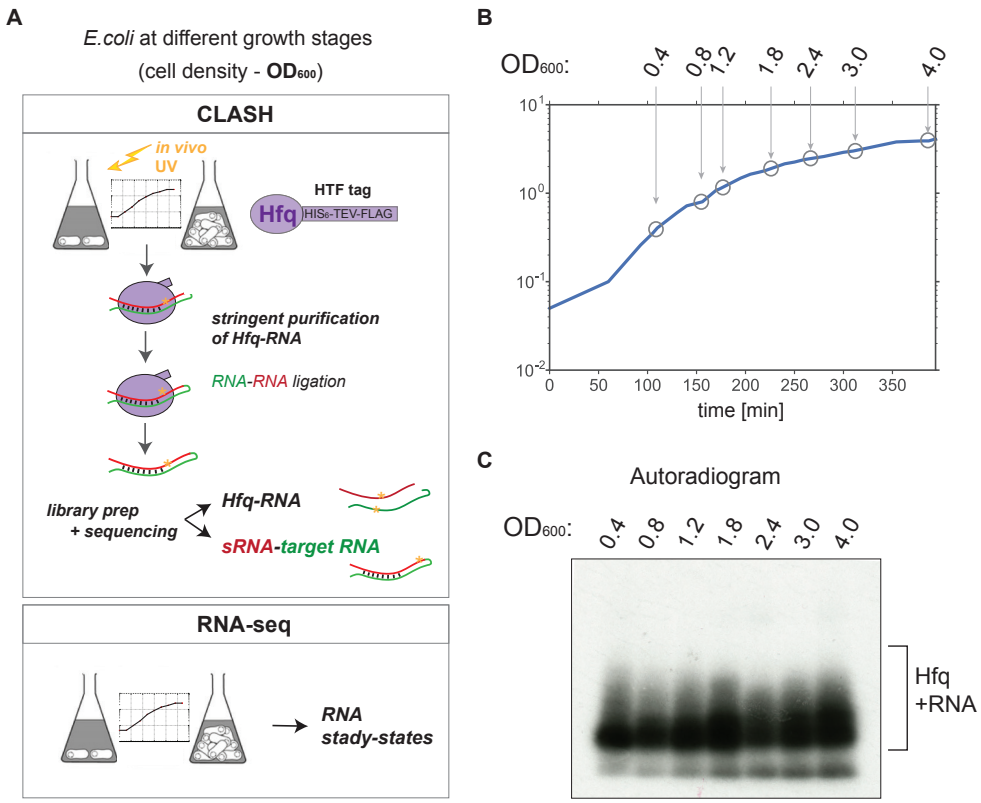

Fig. 3.1: Experimental set-up of the Hfq CLASH experiment at different growth conditions in *E.coli*. (A) Overview of the critical experimental steps for obtaining the Hfq CLASH data and RNAseq data. The cells were grown and crosslinked and harvested at different cell-densities, and the tagged Hfq-RNA complexes were purified in stringent conditions. Hfq binds to sRNA-target RNA duplexes, and the RNA ends found in proximity can be ligated together and sequenced after removal of the protein. The single reads can be used to map Hfq-RNA interactions, whereas the chimeric reads can be traced to sRNA-target interactions. (B) The growth curve of the culture used for the Hfq CLASH experiment, with OD_{600} at which cells were crosslinked indicated by circles. The same OD_{600} were sampled for generating RNA-seq data. (C) Autoradiogram showing the Hfq-RNA complexes purified during CLASH for each OD_{600}

The HTF-tagged Hfq allowed purification of Hfq-bound RNAs in highly stringent conditions (Section 1.8). To detect which transcript classes associate most strongly with Hfq at each $OD_{600,}$ I generated distribution plots of the transcripts crosslinked by Hfq, shown as percentage of total cDNAs per transcript class recovered in each growth condition (Fig. 3.2). The cDNAs represent non-duplicate reads (i.e. multiple reads with the same chromosomal mapping coordinates are considered a single cDNA and are counted only once; details in Section 2.15). The sum of all cDNA counts for all transcript classes at each OD_{600} was set to 100%, and the proportion of each transcript class was calculated as a percentage of the total.

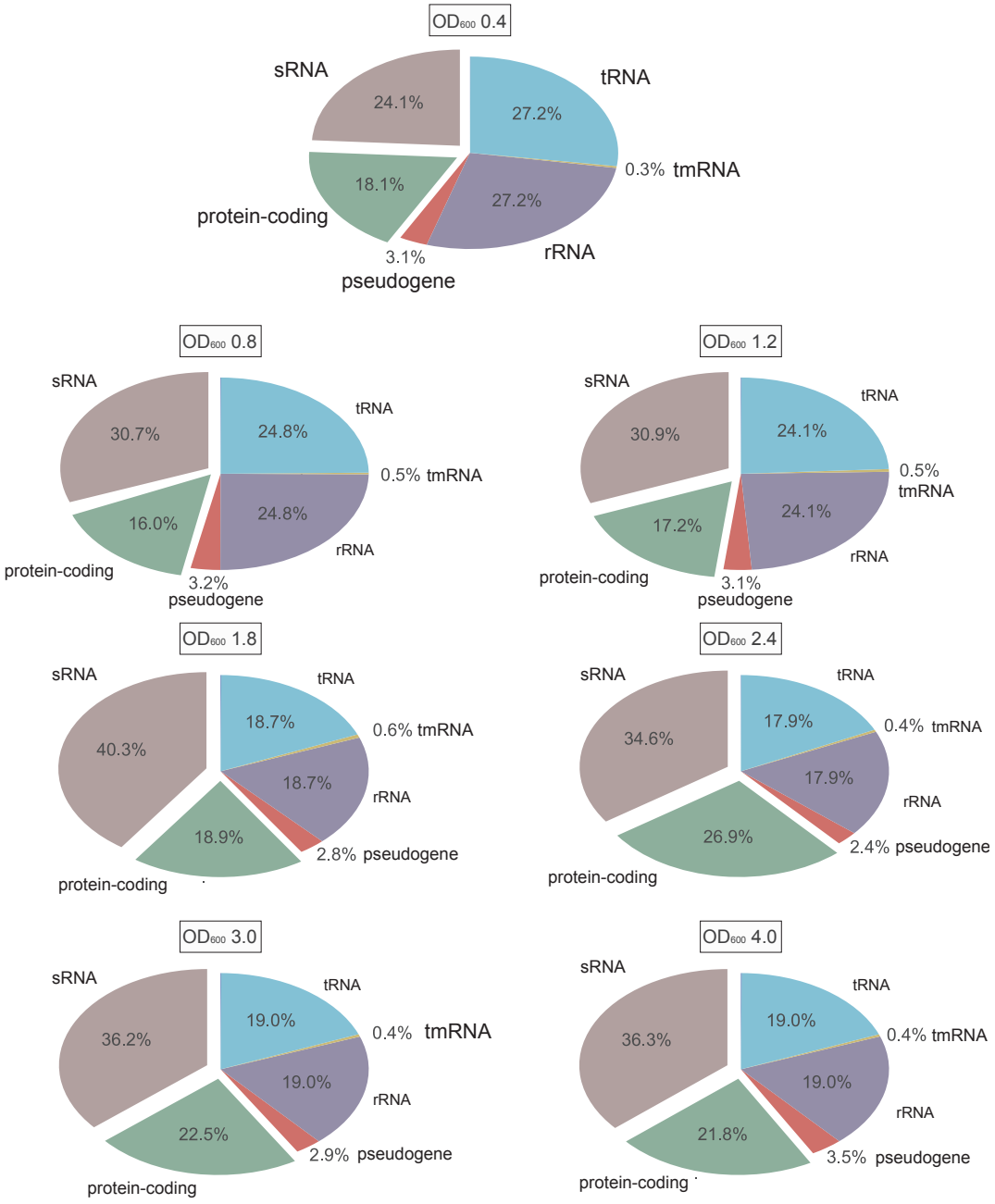

Fig. 3.2: Pie charts showing the distribution of Hfq-crosslinking over genomic transcript classes at different stages of growth in *E. coli* . The growth stage is indicated as OD_{600} above each plot; data are representative of one CLASH experiment per OD_{600}.

The relative distribution of transcript classes among Hfq-crosslinked cDNAs showed that at most optical densities tested, sRNAs were the most enriched transcript class. This was not surprising, given the essential role of Hfq in stabilizing sRNAs and mediating their interaction with target mRNAs (Vogel and Luisi, 2011). The highest proportion of Hfq crosslinking to sRNAs was detected at OD_{600} 1.8 (~40% of all cDNAs),

followed by a drop and stabilization of the percentage at OD_{600} 2.4 - 4 (~35%). The sRNA class was least represented at OD_{600} 0.4.

A relatively high percentage of all cDNAs mapped to tRNAs and rRNAs (on average for all OD_{600} ~20%). However, the Hfq crosslinking to these transcripts also shows a dynamic behaviour: both tRNA and rRNAs appear to associate less with Hfq as the cells grew to higher densities. This is not unexpected, since higher cell densities leads to reduced growth rates and lower ribosome production rates. The proportion of Hfq crosslinking to protein coding transcripts (mRNA) was roughly equal across all growth conditions prior to OD_{600} of 2.4, after which the percentage peaks at OD_{600} 2.4, and as for sRNAs, stabilizes at high OD_{600}. I conclude that the Hfq CLASH data revealed subtle growth-stage-dependent differences in the RNA species associated with Hfq.

3.3. Growth-stage resolved snapshots of Hfq binding to sRNAs

To gain better insight into sRNA association with Hfq, I analysed in more detail the relative Hfq-crosslinking among the known sRNAs (Fig. 3.3). For this, I normalized the cDNAs as TPM (transcripts per million) to correct for sequencing depth and gene length among samples (see section 2.15). Then, I summed the TPM values for all sRNAs at each OD_{600}, and set it to 100%. The “Other” category represents sRNAs that had TPM values lower than 2% of the sum of TPMS. This individual sRNA profiling of Hfq binding further clarifies that in *E. coli* there are cell density-dependent changes in Hfq-sRNA binding, as observed in studies in *Salmonella* (Chao et al., 2012). Many sRNAs were bound by Hfq in all growth-stages. Several sRNAs associate with Hfq more strongly than others, in a condition-specific manner. For example, Spf (Spot42 sRNA) is the most highly cross-linked sRNA at OD_{600} 0.4, but shows a dramatic decrease in association with Hfq at later time-points. This binding profile is consistent with the Spf documented role in catabolite repression (Beisel and Storz, 2011), which, during growth in LB, becomes less pronounced at high OD_{600} (Sezonov et al., 2007). GcvB, an sRNA involved in amino acid metabolism, shows a decrease with OD_{600}. Both examples confirm the growth-dependency of sRNA association

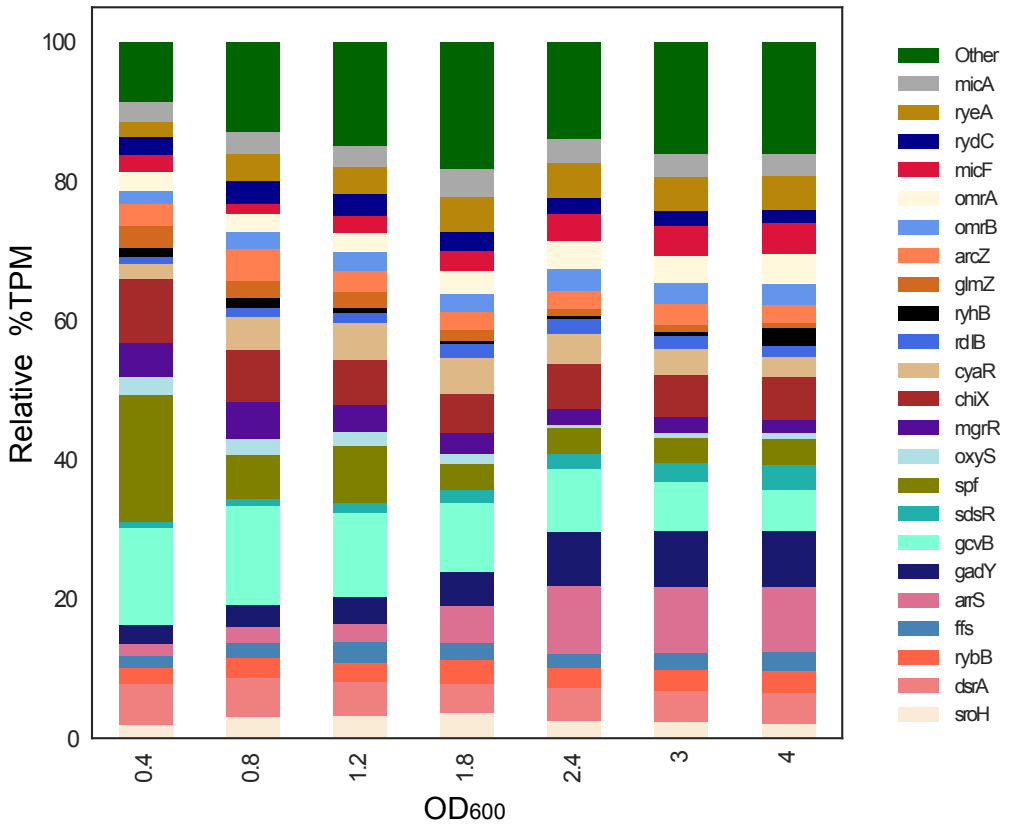

Fig. 3.3: Dynamics of Hfq-sRNA binding during growth in LB in *E.coli.* Stacked bar charts showing the relative Hfq-crosslinking to sRNAs at different stages of growth in *E.coli* as percentage of total TPM (transcripts per million). The growth stage is indicated as OD_{600} units below each bar; data are representative of one biological replicate for each OD_{600}; the 'Other' category includes the sRNAs for which cDNAs %TPM was below an arbitrary threshold of 2%.

with Hfq reported in *Salmonella* (Chao and Vogel, 2012). GlmZ is another sRNA that shows a decreasing trend in Hfq association, a pattern consistent with its expression profile (Argaman et al., 2001). Conversely, some sRNAs show increased binding to Hfq with OD_{600} (ArrS, SdsR; OmrA/B - slight increase). A gradual, yet large increase in Hfq binding to ArrS was observed - consistent with the fact that its transcription is activated in stationary phase and is σ^S-dependent (Aiso et al., 2014). However, ArrS is a *cis*-encoded sRNA, a class typically characterized by a lack of requirement of Hfq for function, so it is intriguing that ArrS crosslinks so well to Hfq. Some sRNAs show minor fluctuations or relatively constant association with Hfq (ChiX, DsrA, MicA and RybB). Only several sRNAs showed an increase in the percentage of binding to Hfq, followed by a decrease. One example is CyaR, whose transcription is dependent on cAMP levels (De Lay ad Gottesman, 2009). Noteworthy, this type of analysis of

Hfq association is not necessarily a representation of growth-dependent Hfq binding to individual sRNAs, or not even their expression levels - rather, it indicates which sRNAs associate more strongly with Hfq relative to others *in vivo*, possibly providing a snapshot of the dynamics of competition between sRNAs for Hfq. In many cases, however, these effects are influenced by the expression level of the sRNAs.

To analyze Hfq binding to individual sRNAs and to understand how this compares to their RNA levels, I generated 10 clusters that represent different patterns of growth-dependent Hfq binding to all transcripts. The clusters were generated by K-means clustering using TPM normalized Hfq CLASH data (Fig. 3.5). Briefly, for data normalization, for each gene I calculated the $\log_2$-fold change in TPM values relative to the first OD_{600} = 0.4 time-point. Consequently, the starting point for all clusters is set at 0 (= log2 [Fold change OD_{600} 0.4 relative to OD_{600} 0.4]). Similarly, I generated 10 clusters representing the patterns in total RNA levels using the RNA-seq data.

Fig. 3.4 zooms in the sRNAs belonging to each cluster. Cluster 0 (Fig. 3.5 and Fig. 3.4, left) contains gene whose average profile shows no dramatic changes in relative expression with OD_{600} (less than 2-fold differences). A notable example belonging to this cluster is ArcZ, that shows very little change in Hfq binding relative to OD_{600} 0.4, consistent with the data in Fig. 3.3. Interestingly, this pattern of Hfq binding did not always correlate with changes in the sRNA levels. With respect to ArcZ expression, this sRNA belongs to the RNA steady-state cluster 8, a cluster that contains sRNAs whose expression levels show an increase in RNA levels after OD_{600} 2.4. A similar case is DsrA. MicA (Hfq binding cluster 0) displays the highest fold-change relative to OD_{600} 0.4, at OD_{600} 2.4, but the expression level fold-change steadily increases with OD_{600} and peaks at OD_{600} 4 (RNA steady-state cluster 3). Another example is Spf, that belongs to Hfq binding cluster 8. This sRNA shows a decrease in Hfq cross-linking relative to OD_{600} 0.4 (e.g. cross-linking at OD_{600} 4.0 is reduced relative to OD_{600} 0.4), but the RNA levels at OD_{600} of 4.0 are comparable to Spf levels at OD_{600} 0.4 (RNA steady-state cluster 5).

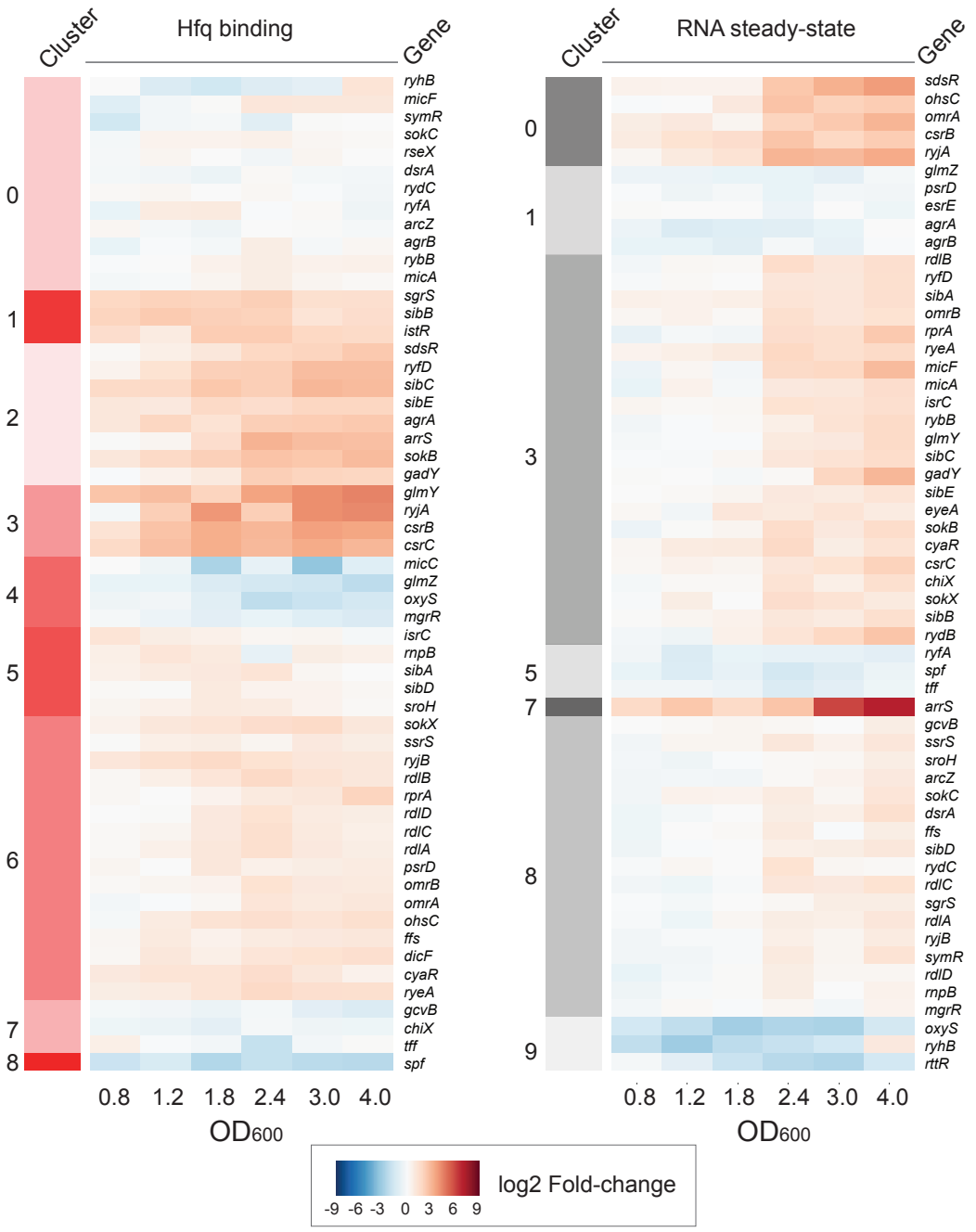

Fig. 3.4: Growth-dependent sRNA expression and Hfq binding in *E.coli*. Heatmaps showing the clustered, normalized Hfq-crosslinking (Left; using the CLASH data) and RNA steady-states (using the RNAseq data) of sRNAs relative to the starting point (OD_{600} 0.4). The cluster assignment is indicated at the left of each plot. The clusters were generated by k-means clustering as follows: The TPM (transcripts per million) values of genes with minimum 5 TPM at each OD_{600} for all genes (including sRNAs and protein-coding genes), were divided by the TPM values corresponding to OD_{600} 0.4. The resulting ratios were log_2-normalized, yielding the log_2 Fold-change that was used as input for the K-means algorithm, which calculated 10 clusters. A blue shade indicates a low fold-change, whereas a dark orange shade indicates a high fold-change relative to OD_{600} 0.4.The growth stage is indicated as OD_{600} units below each heatmap; data are representative of one biological replicate for each OD_{600}.

The Hfq-binding cluster 3 average profile is characterized by a dramatic increase in Hfq cross-linking relative to OD_{600} 0.4 (more than 8-fold) at OD_{600} 1.8, followed by a stabilization in the levels of Hfq binding. Interestingly, CsrB and CsrC sRNAs belong to this cluster, and overall, they seem to be among the sRNAs that show the largest increase in Hfq binding relative to OD_{600} 0.4 (8-10-fold). Whereas for CsrB, this change in Hfq cross-linking roughly recapitulates the sRNA steady-state levels (RNA steady state cluster 2), CsrC shows a slightly different expression profile, belonging to RNA steady state cluster 3.

ArrS (Hfq-binding cluster 2 and RNA steady-state cluster 7) is an example that illustrates some consistency in the patterns of expression and Hfq binding - i.e. the increased Hfq cross-linking can be partly explained by the high expression and fold-change increase in the expression of this sRNA relative to OD_{600} 0.4. However, the fold-change in RNA levels is more pronounced than the increase in Hfq binding. I speculate that this difference illustrates its role as a dual sRNA - a *cis*-encoded sRNA that can also act as a *trans*-encoded sRNA. The pool of ArrS that do not bind Hfq may act as *cis*-encoded sRNAs. Corroborating on the latter hypothesis, hybrids that support ArrS interaction with mRNAs (and even sRNAs) were uncovered in the CLASH data (Section 3.6, 3.8). An example that illustrates a relatively good correlation between changes in RNA expression and Hfq binding is CyaR (Hfq binding cluster 6 vs. RNA steady-state cluster 3).

Overall, the comparison of the two types of sRNA clusters shows that in many cases, the increase in Hfq binding to sRNAs does not always recapitulate changes in sRNA expression levels. I investigated the correlation more closely in Section 3. 5.

3.4. Growth-stage resolved snapshots of Hfq binding to mRNAs

Following the binning of all transcripts with similar Hfq binding profiles into 10 clusters (Section 3.3 and Fig. 3.5), I performed a gene ontology (GO) term and KEGG pathways enrichment analysis of the protein-coding genes belonging to each cluster (Fig. 3.6).

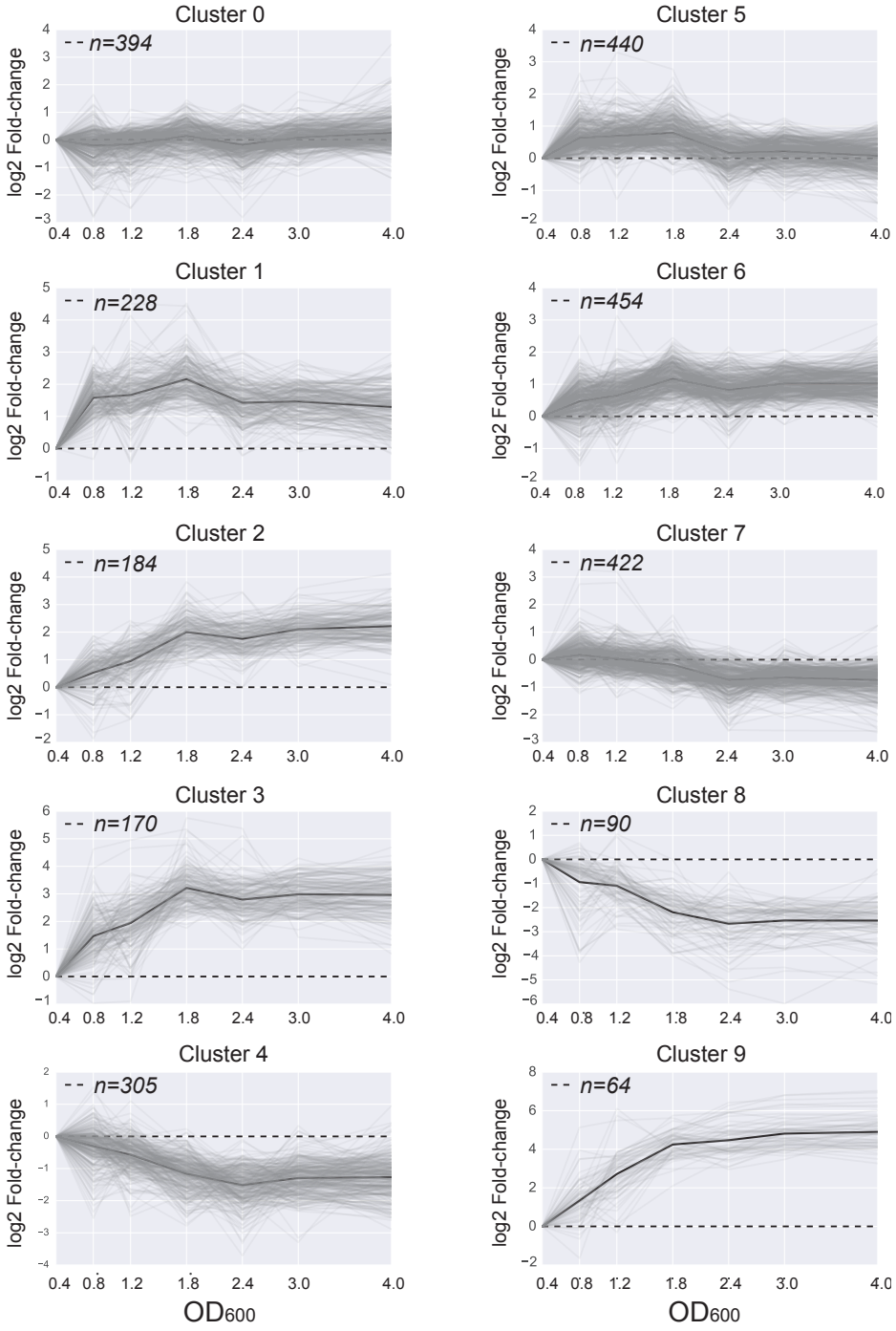

Fig. 3.5: Profiles of growth-dependent Hfq binding in *E.coli*. Hfq crosslinking profiles for protein-coding genes from one CLASH dataset were generated by k-means clustering using python packages (Sklearn.cluster - Kmeans). Profiles of individual genes (grey lines) were used to generate the average profile for each cluster (black lines). The number of genes belonging to each cluster (n) is indicated at the top left of each graph; the y-axis indicates log_2-fold change of TPM of each sampled OD_{600} relative to time-point 0 (OD_{600} 0.4); the y-axis indicates the analyzed growth-stages (OD_{600}). Data are representative of one biological replicate for each OD_{600}.

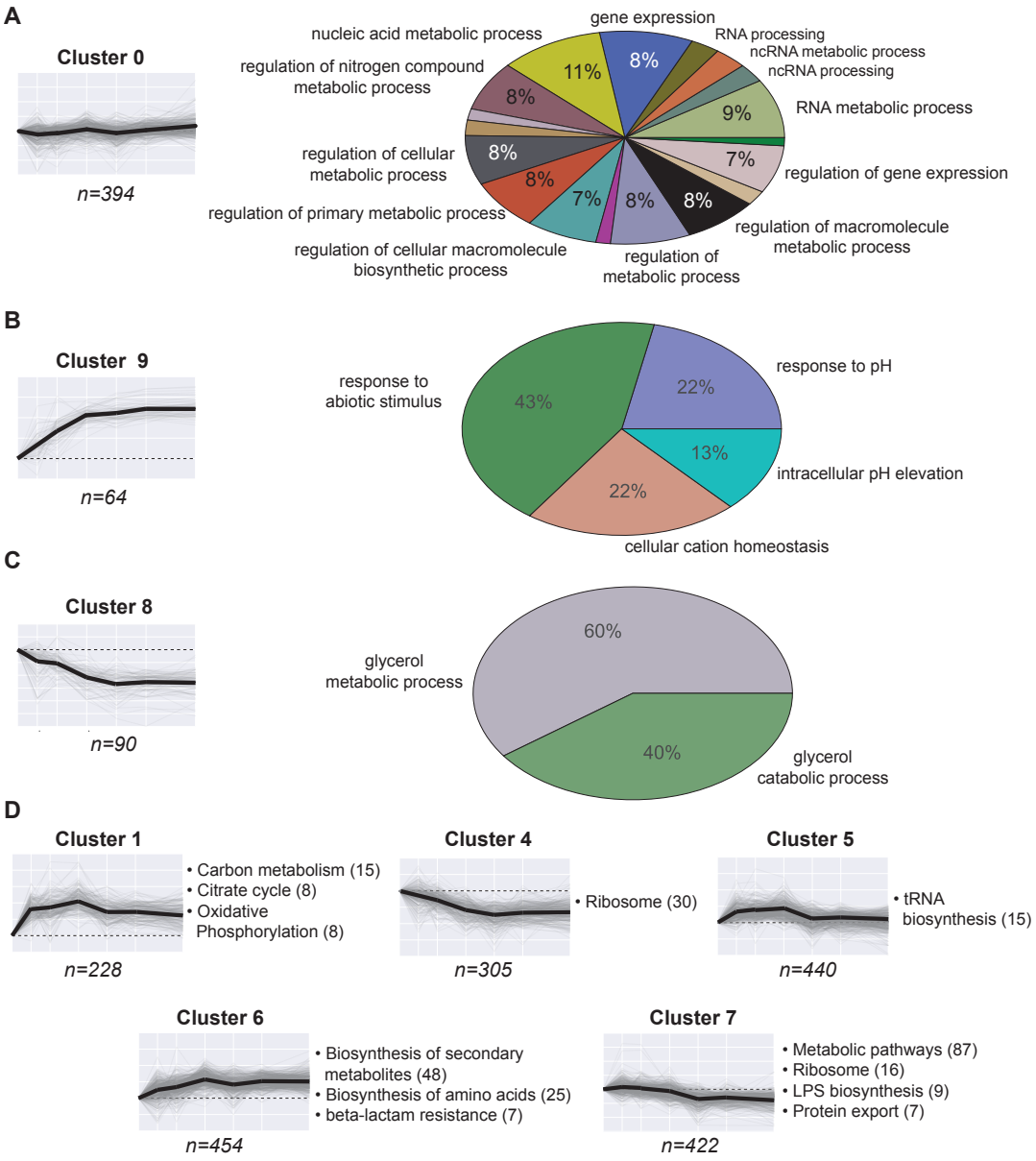

Fig. 3.6: The biological roles of protein-coding genes belonging to each Hfq-binding profile cluster indicated by GO (gene ontology) terms (A-C) and KEGG-pathways (D) enrichment analysis.. Hfq-crosslinking profiles were generated by k-means clustering as in Fig. 3.5; the dark thick line indicates the average profile for each cluster.; the enrichment analysis was performed using STRING (Szklarczyk et al., 2015). (A-C) on the right of each graph, the pie charts show the distribution of enriched GO (FDR <0.01 in A; FDR < 0.03 in B-C). (D) Enriched KEGG pathways are indicated at the right of each graph; the number of genes in each term are indicated in brackets.

Cluster 0, which includes genes that show no dramatic fold-changes in Hfq binding relative to OD_{600} 0.4 (Fig. 3.5 and Fig. 3.6 A), was enriched in many biological process GO terms - however, none of these GO terms were statistically significantly over-represented, and all seem to be processes necessary for the general maintenance

of the cell (e.g. regulation of primary metabolic process, RNA processing and metabolism, regulation of translation). These observations hint to the fact that Hfq may 'constitutively' regulate these processes throughout growth.

Two of the most interesting clusters are 8 and 9 (Fig. 3.5). Although they include a small number of genes (90 and 64, respectively), they do show clearly defined expression profiles, with few outliers. Cluster 9 (Fig. 3.6B) shows a steady increase in Hfq binding with cell-density (higher slope until OD_{600} 1.8) and is enriched in genes encoding proteins that respond to acidic-condition induced stress. Conversely, cluster 8 (Fig. 3.6C) is characterized by a steady decrease in Hfq binding fold-changes relative to OD_{600} 0.4 and is enriched in genes involved in glycerol metabolism.

The other clusters that showed GO-term enrichment covered a wide spectrum of biological functions - for simplicity, the enrichment of the KEGG pathways is discussed here (Fig. 3.6 D). Clusters 4 and 7, that both include genes with profiles that show slight decrease in Hfq binding, are both involved in ribosome pathways, whereas cluster 6, that shows a slight increase in Hfq binding during growth, includes genes with roles in beta-lactam resistance.

To understand what role Hfq may have in the regulation of the genes within clusters, I focused on clusters 8 and 9 (Fig. 3.7). First, I analysed the RNA steady-state levels of the genes in clusters 8 and 9. The average profile of the RNA expression levels is roughly similar to the Hfq-binding average profile; however, many genes deviated from the average profile of the RNA levels (Fig. 3.7A). I cannot conclude whether reduced Hfq binding to these genes is a consequence of reduced RNA steady states, or vice-versa. In the absence of transcriptional activity data for these genes, it is very difficult to make predictions on the effect of Hfq binding based on RNAseq and Hfq CLASH data alone. However, analysing the distribution of Hfq binding to these genes may help uncover the mechanism employed by Hfq to regulate these genes. To make the distribution plots, I selected the genes from each cluster and divided their length in 100 bins, then the hit density in each bin was calculated (Fig. 3.7 B) - for each

cluster, I generated distribution plots for the OD_{600} with maximal fold-change in Hfq-binding relative to OD_{600} 0.4 (ie OD_{600} 4.0 for cluster 9 and OD_{600} 0.8 for cluster 8). Hfq binding to 5'-UTRs of mRNAs may preclude the 30S ribosomal subunit association with the SD sequence (Section 1.4; Bouvier et al., 2008) or activate translation, whereas interactions further downstream (CDS and 3'-UTR) may impede translation by RNase E recruitment and degradation of the transcript (Pfeiffer et al., 2009; Bandyra et al., 2012). Although Hfq-binding was somewhat enriched at the 5'-ends of transcripts in both clusters, clearly, Hfq binding to the coding sequences and 3' ends (including 3' UTRs) of the mRNAs occurred. Thus most likely these transcripts are not regulated by Hfq in similar ways. A recently described mechanism (Sedlyarova et al., 2016) of sRNA action involves prevention of premature termination by sRNAs by preventing Rho recruitment to 5'-UTRs, and the authors postulated that the phenomenon could be widespread. To test whether this could be the case for the genes in clusters 8 and 9, I calculated the average length of the 5'-UTRs of these genes: cluster 8 had an average 5'-UTR of 77 nucleotides, whereas cluster 9 5'-UTRs had an average length

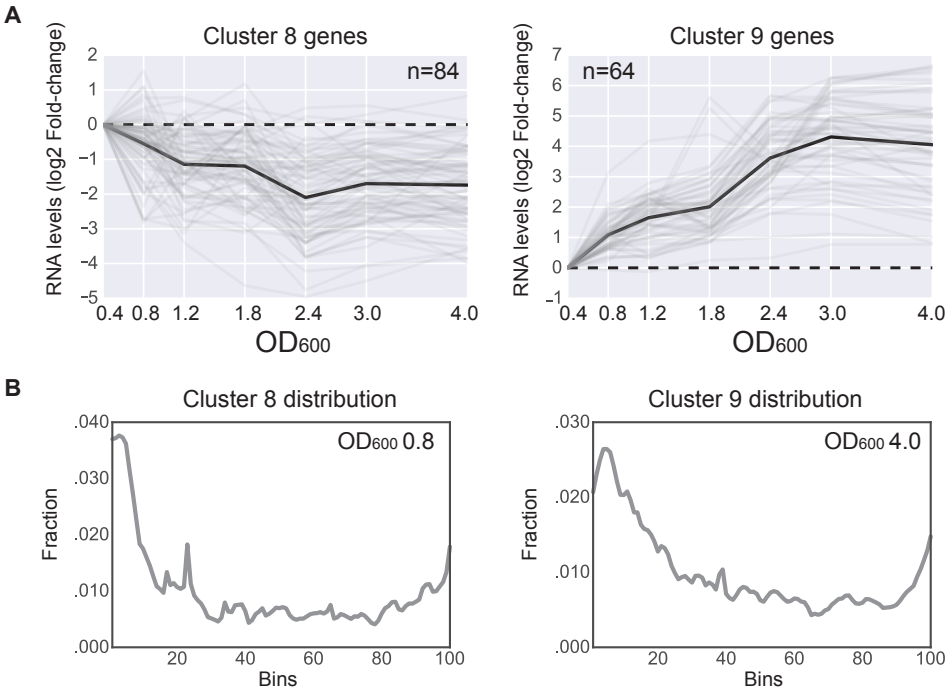

Fig. 3.7: Characterization of Cluster 8 and 9 transcripts. (A) The RNA levels of the genes from Clusters 8 and 9. The black line indicates the average profile of expression, but is not representative for all genes, i.e. the genes are not part of a defined cluster describing RNA expression. (B) Transcriptome-wide coverage plots of Hfq binding. The gene lengths were normalized by dividing them into 100 bins (x-axis), and the fraction of hits in each bin was calculated (y-axis).

of 50 nucleotides. Given that Rho typically requires at least 80 nucleotides of relatively unfolded RNA sequences to bind (Sedlyarova et al., 2016), it is unlikely that many the transcripts in these clusters are transcriptionally attenuated in the absence of sRNA-mediated regulation.

I am now in the process of looking for chimeras that mapped to 5' UTRs of genes in these clusters. I expect that this will help me to better predict how Hfq may regulate the expression of these genes.

3.5. Comparison of transcriptome-wide changes in Hfq-binding to changes in RNA levels during growth

The analysis of Hfq-binding profiles and comparison of the RNA level fold-changes of the corresponding genes for both sRNAs (Fig. 3.4) and mRNAs (Fig. 3.5 vs Fig. 3.7 A) revealed that in some cases, the changes in Hfq binding appear to not be correlated with the changes in RNA levels. To dissect how well the Hfq CLASH data correlates with changes in RNA steady-states, and to check if the correlation globally changes during growth, I compared the changes in Hfq binding to the changes in total RNA levels for all transcripts (RNA-seq data), at each $OD_{600.}$ As for the cluster analyses, I normalized the data by calculating the $\log_2$ (TPM Fold-change relative to OD_{600} 0.4). Only the genes that were present in both datasets were used to build the correlation. A quick overview of the plots showing the changes in RNA levels versus changes in Hfq cross-linking, and the Pearson correlation coefficient associated with each (r) (Fig. 3.8) , shows that the correlation between Hfq-binding fold-changes and RNA steady-state fold-changes is quite poor at lower cell density (OD_{600} 0.8: r = 0.38), but gradually becomes positive as the cells approach stationary phase (OD_{600} 3.0 r =0.75). These dynamics are intriguing, but I cannot fully explain this observation.

A starting point, however, is to analyse the outliers. In the following analyses, I defined highly-crosslinked RNAs and highly-expressed RNAs as those that show a $\log_2$ fold-change relative to OD_{600} 0.4 higher than 2 (i.e at least a four-fold change), an

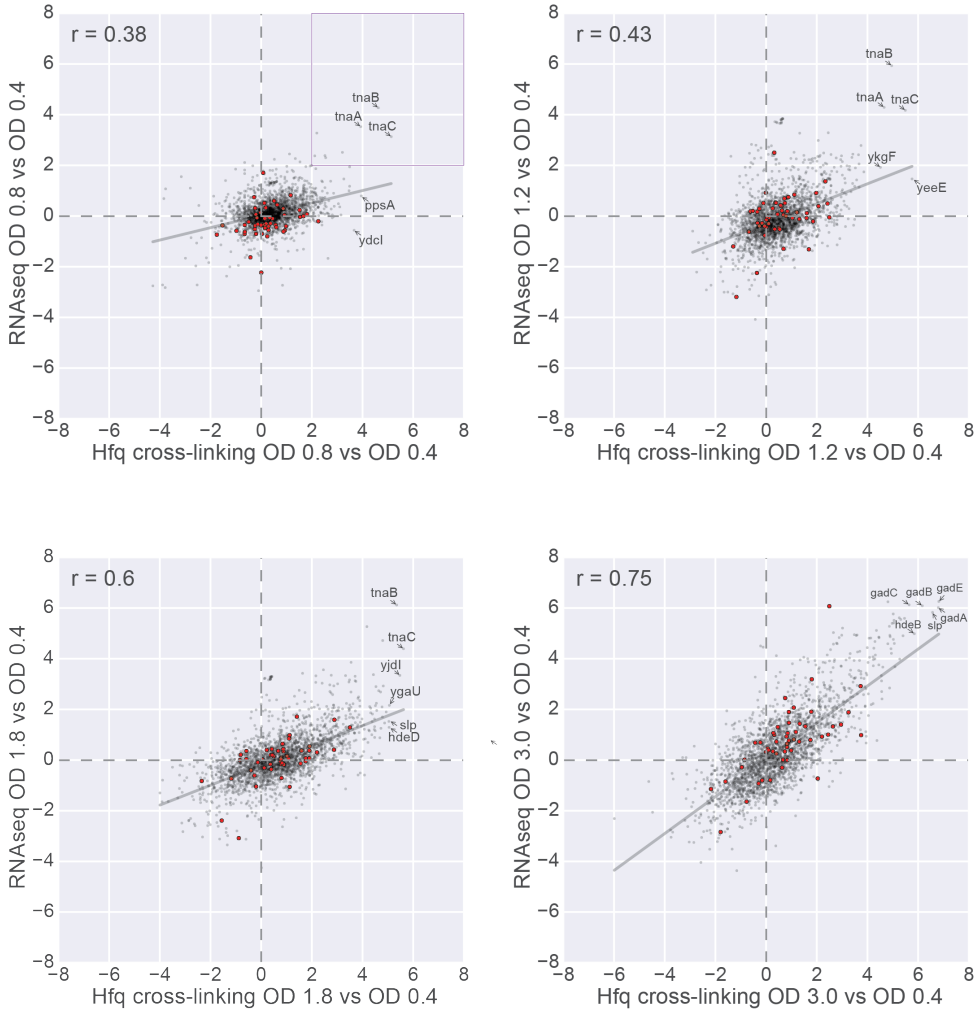

Fig. 3.8: Comparison of changes in Hfq binding to changes in RNA levels. Scatter plots comparing the Hfq binding (x-axis) to RNA levels (y-axis) for the indicated OD_{600}, for all transcripts. Small RNAs are indicated in red. For each $OD_{600,}$ from 0.8 to 4.0, the TPM values at each time-point were divided by the OD_{600} 0.4 data, and the resulting ratios were log_2-normalized. Data are representative of one biological replicate for each $OD_{600.}$ The top 5 highest-crosslinked transcripts are annotated. The Pearson coefficient (r), calculated for all transcripts is indicated on each plot.

arbitrary but very stringent threshold. First, I selected those mRNAs that are highly-crosslinked for three representative conditions: OD_{600} 0.8 for exponential phase of growth, OD_{600} 1.8 for transition between exponential phase and stationary phase of growth, and OD_{600} 3.0 for entry into stationary phase. I intersected the sets to see which mRNAs are highly bound to Hfq in all growth conditions, and to see which of the mRNAs crosslink in a condition-specific manner (Fig. 3.9 A). There were 53 mRNAs most-strongly associated with Hfq in all conditions, and these were enriched in GO-

terms related to amino acid metabolism (14 genes). The biggest overlap was between mRNAs highly-crosslinked at OD_{600} 1.8 and OD_{600} 3.0, suggesting that these mRNAs already start being regulated by Hfq at the transitions during growth stages and the regulation is maintained as the cells reach stationary phase. A GO-term analysis of the transcripts highly-bound by Hfq at OD_{600} 3.0 revealed that most of these RNAs are involved in responses to stress (53 genes), which is not unexpected as cells undergo major changes in gene expression while they transition to stationary phase, a stage of growth characterised by increased resilience to stress. Further analyses of the transcripts that are strongly bound to Hfq in only one condition will be performed and I expect them to reveal more insights into the growth stage-specific association of mRNAs to Hfq.

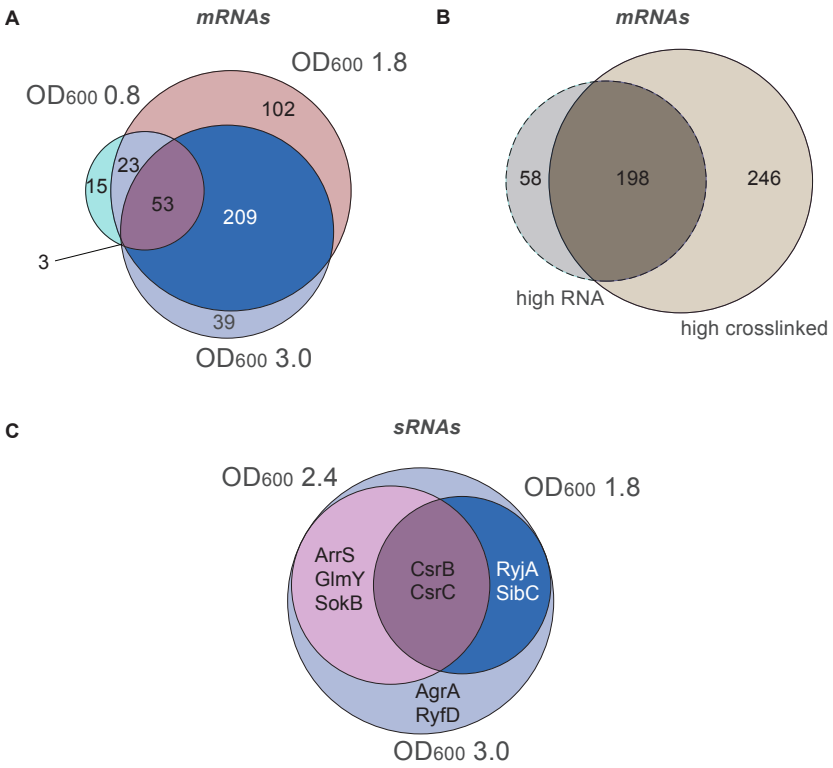

Fig. 3.9: Overview of highly-cross-linked RNAs. The threshold for high cross-linking was set at $\log_2$ Fold-change >2 (four-fold change) relative to OD_{600} 0.4 data. (A) Venn diagram showing the intersection between highly-crosslinked mRNAs in three different growth conditions. (OD_{600}). (B) Venn diagram that shows the intersection between highly expressed mRNAs and highly-cross-linked mRNAs for combined/reunited datasets at all $OD_{600.}$ The threshold for high expression was set at $\log_2$ Fold-change >2 (four-fold change). (C) Venn diagram showing the intersection between highly-crosslinked sRNAs in three different growth conditions. The names of the sRNAs are indicated on their corresponding set.

One simple way to spot outliers among the highly Hfq-crosslinked mRNAs is to check if the high crosslinking is correlated with high expression, and investigate those RNAs. To get an initial overview of this, I reunited the highly-crosslinked mRNA genes of all three conditions (OD_{600} 0.8, 1.8 and 3.0) and compared the resulting set of genes with a similarly generated set containing the highly-expressed mRNA genes in the corresponding growth stages (Fig. 3.9 B). Although roughly 200 of these RNAs are both highly expressed and cross-linked (note: the dynamics could still be different!), 250 more mRNAs are more highly cross-linked than expressed. Further approaches involve modifying the stringency of the parameters (four-fold may be too stringent), and investigate the following: which RNAs are highly cross-linked but show no change in RNA expression level; which individual mRNAs are changing their correlation throughout growth. I am currently exploring these avenues and for interesting mRNAs I will check if there are any RNA-RNA interactions that could partly explain the changes in correlation.

3.6. Changes in Hfq-binding to sRNAs may reveal unconventional roles for some sRNAs

I next asked how well the Hfq-crosslinking sRNA data correlated with changes in sRNA steady-state levels (Fig. 3.10). The trend of correlation improvement with OD_{600} is preserved - however, it appears that globally, the changes in sRNA crosslinking relative to OD_{600} 0.4 are more poorly correlated to sRNA levels, compared to the whole transcriptome (Fig. 3.8). Some sRNAs, such as Tff and SokB are close to the trendline in all growth conditions (Fig. 3.10). However, other sRNAs deviate from the trend. The most striking example is ArrS, an RNA that is very highly-expressed throughout growth relative to OD_{600} 0.4, but the increase in Hfq cross-linking is comparatively low (except OD_{600} 2.4) (also see Section 3.3). Conversely, AgrA is a sRNA that shows a decrease in expression relative to OD_{600} 0.4 (although expression starts increasing after OD_{600} 4.0), but the Hfq binding increases (especially from OD_{600} 1.8 to OD_{600} 2.4). What these two sRNAs have in common is that they are typically known as cis-encoded sRNAs.

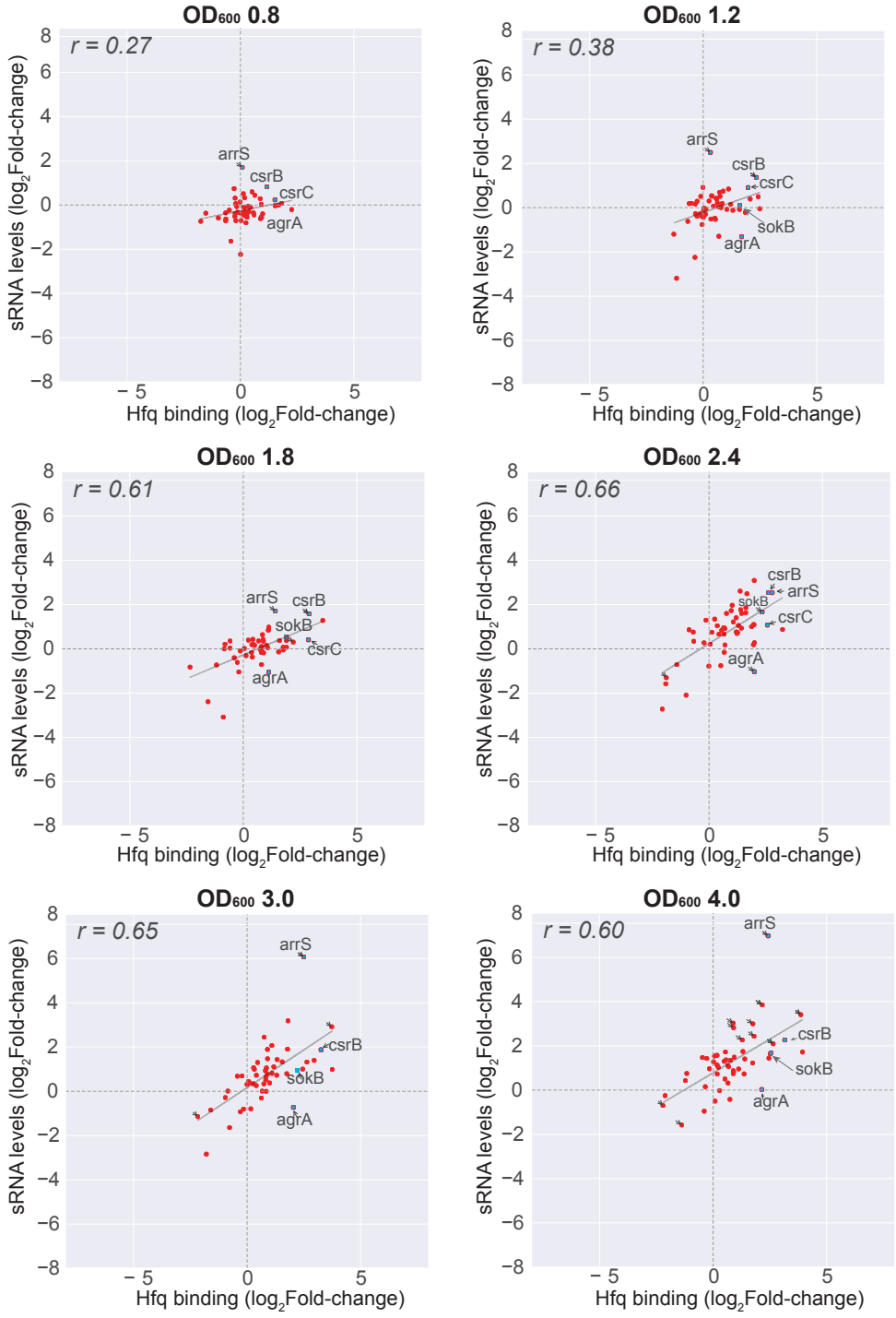

Fig. 3.10: Comparison of changes in Hfq binding to changes in sRNA levels. Scatter plots comparing the Hfq binding (x-axis) to aRNA levels (y-axis) for the indicated OD_{600}. For each $OD_{600,}$ from 0.8 to 4.0, the TPM values at each time-point were divided by the OD_{600} 0.4 data, and the resulting ratios were log2-normalized. Data are representative of one biological replicate for each OD_{600}. The sRNAs discussed in the text are annotated in blue. The Pearson coefficient (r) calculated for sRNAs only, is indicated on each plot.

Further, I investigated which sRNAs are very highly-crosslinked to Hfq relative to OD_{600} 0.4, in three conditions that encompass the transition to and entry into stationary

phase (OD_{600} 2.4 to 4.0) (Fig. 3.9 C). Interestingly, the sRNAs with high-fold changes in cross-linking relative to OD_{600} 0.4 in all tested conditions are CsrB and CsrC, sRNAs conventionally known for their Hfq-independent function in sequestering the CsrA protein (Section 1.4; Romeo et al., 1998).

To dissect the Hfq binding to CsrB, I determined the read density profile across the CsrB transcript, defined by the total number of times each nucleotide within reads was mapped to CsrB. Additionally, since deletions within sequences may be indicative of the actual site of cross-linking (Granneman et al., 2009), I counted the deletions occurring at each nucleotide position. I performed these analyses for the conditions in which CsrB was most highly bound to Hfq relative to OD_{600} 0.4, namely at OD_{600} 3.0 and 4.0. These two conditions can also be viewed as independent replicates to support crosslinking sites, i.e. finding the same position mutated at both optical densities could confirm that the deletions are not sequencing mistakes (Fig. 3.11A). The positions of the deletions were superimposed on the CsrB predicted secondary structure (Fig. 3.11 B). CsrB has ~18 CsrA binding sites (Liu et al., 1997). Although not all putative sRNA binding sites contained deletions indicating Hfq binding, at least seven crosslinked to Hfq, and often the crosslinking sites occurred within or near the GGA binding motif found in the loops. This partial overlap with CsrA binding (Holmqvist et al., 2016) may either implicate a role for Hfq in masking CsrA binding sites, or Hfq may bind to CsrB to mediate CsrB-mRNA base-pairing interactions.

3.7. Maps of Hfq binding sites across protein-coding genes

The analyses presented so far mainly focused on the quantitative aspect of Hfq binding to sRNAs and mRNAs. This section will explore the distribution of Hfq binding sites across all protein-coding genes.

To identify where Hfq preferentially binds to mRNAs, I generated heat maps that show the distribution of Hfq binding sites across all the transcript coding sequences flanked by UTRs of a set length of 200 nt (Fig. 3.12A). The number of genes and the patterns

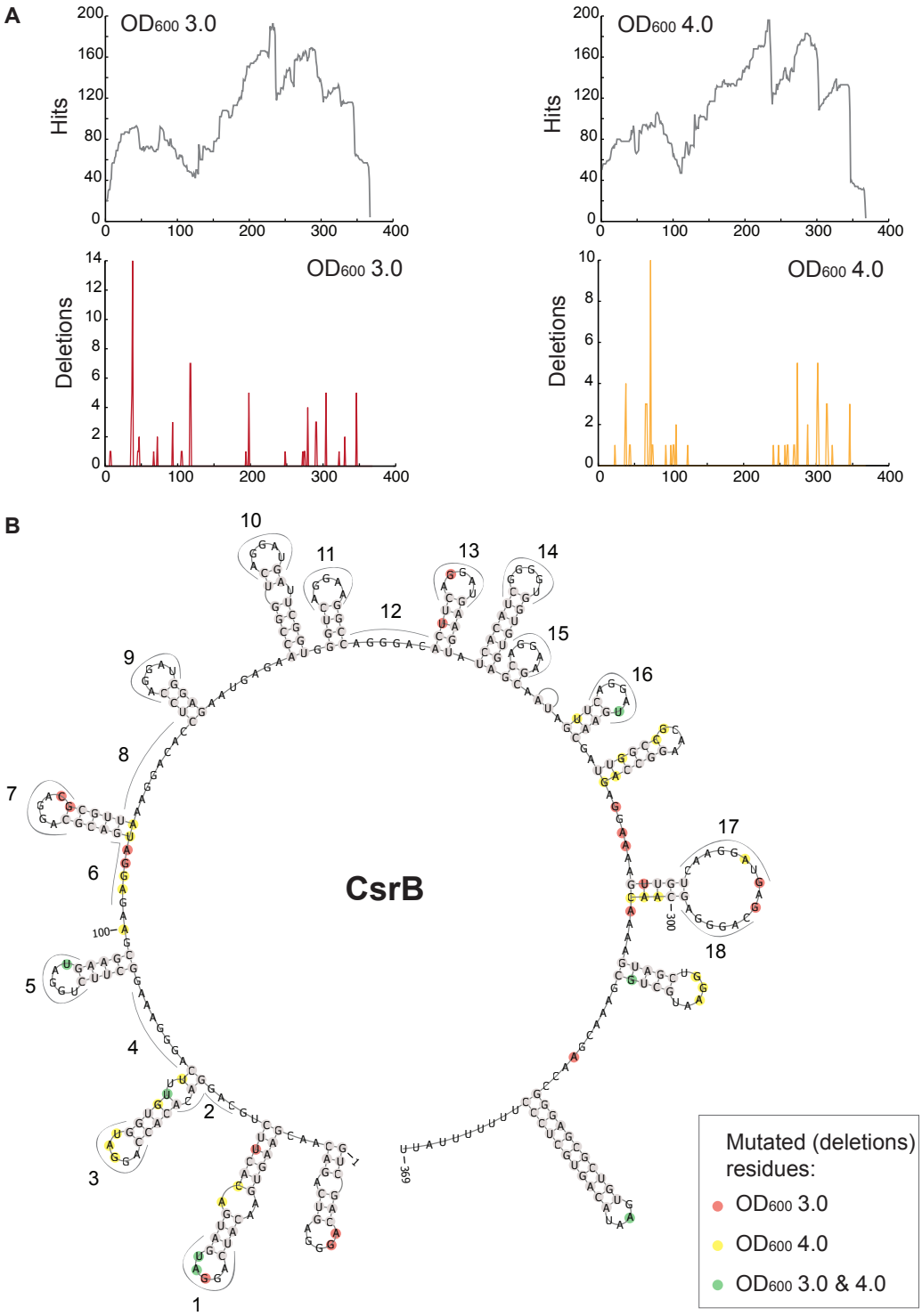

Fig. 3.11: Analysis of Hfq crosslinking to CsrB sRNA. (A) The lines in grey show hit density at OD_{600} 3 (left) and 4 (right). The y-axis corresponds to the total number of times (hits) each nucleotide within an RNA mapped to the full CsrB sequence. The x-axis indicates the nucleotide position. The lines in red and yellow show the deletions at each nucleotide position of CsrB, indicative of sites of Hfq cross-linking. (B) CsrB secondary structure drawn in Pseudoviewer 3 (Byun and Han, 2009), based on published secondary structure predictions, with numbers indicating putative CsrA binding sites (Liu et al., 1997). The nucleotide positions in which deletions were found are indicated with circles: green fill indicates the residues that were mutated at both OD_{600} 3 and 4; red fill indicates deletion sites at OD600 4, and yellow fill indicates deletion sites at OD_{600} 3 only.

of Hfq binding site distribution did not differ dramatically among late exponential phase (OD_{600} 0.8), transition to stationary phase (OD_{600} 1.8) and entry into stationary phase (OD_{600} 3.0). Since the 5'-UTR of one gene may overlap with the 3'-UTR of a nearby gene, thus resulting in possible redundant Hfq binding, I have removed those genes with overlapping UTRs from the analysis. Moreover, I analysed only those transcripts with enriched Hfq binding (FDR intervals, minimum 20 nt length) and used the UTR-lengths from the Rockhopper GTF annotation file (Fig. 3.12 B). Even with this more stringent approach, the profile of binding sites across the protein-coding transcriptome is preserved. Although Hfq binding was found in all mRNA features (5'-UTRs, coding sequences, and 3'-UTRs) (Fig. 3.12 A), most Hfq sites mapped to 5'-UTR and 3'-UTR mRNA regions. The enrichment of Hfq binding to the 5'-UTRs of mRNAs is consistent with the many regulatory roles Hfq performs at this site, many involving sRNA base-pairing (Section 1.4), but also direct Hfq-binding without an sRNA to inhibit translation (Ellis et al., 2015). Similarly, at 3'-UTRs, Hfq may have multiple roles, such as sRNA-mediated stabilisation of transcripts (e.g. gadY; Frohlich and Vogel, 2009; Papenfort and Vanderpool, 2015), promotion of degradation (Morita et al, 2005). However, an emerging role of Hfq that would explain 3'-UTR binding enrichment is to stabilize sRNAs that overlap the ORFs of mRNAs (Section 1.7).

Next, I used the FDR intervals to search for a consensus motif for Hfq-binding to mRNAs (Fig. 3.12C). The most enriched K-mer included polyU stretches, that resemble the polyU tracts characteristic to Rho-independent terminators (Wilson and von Hippel, 1995). Many bacterial transcripts terminate with such terminators, and the finding this motif is consistent with the observed Hfq-enrichment at the 3'-UTRs of mRNAs. These findings confirm the motif uncovered in CLIP-seq studies in *Salmonella* (Holmqvist et al., 2016), but do not cross-validate the motif proposed from Hfq CRAC experiments in enterohemorrhagic *E.coli* (Tree et al., 2014).

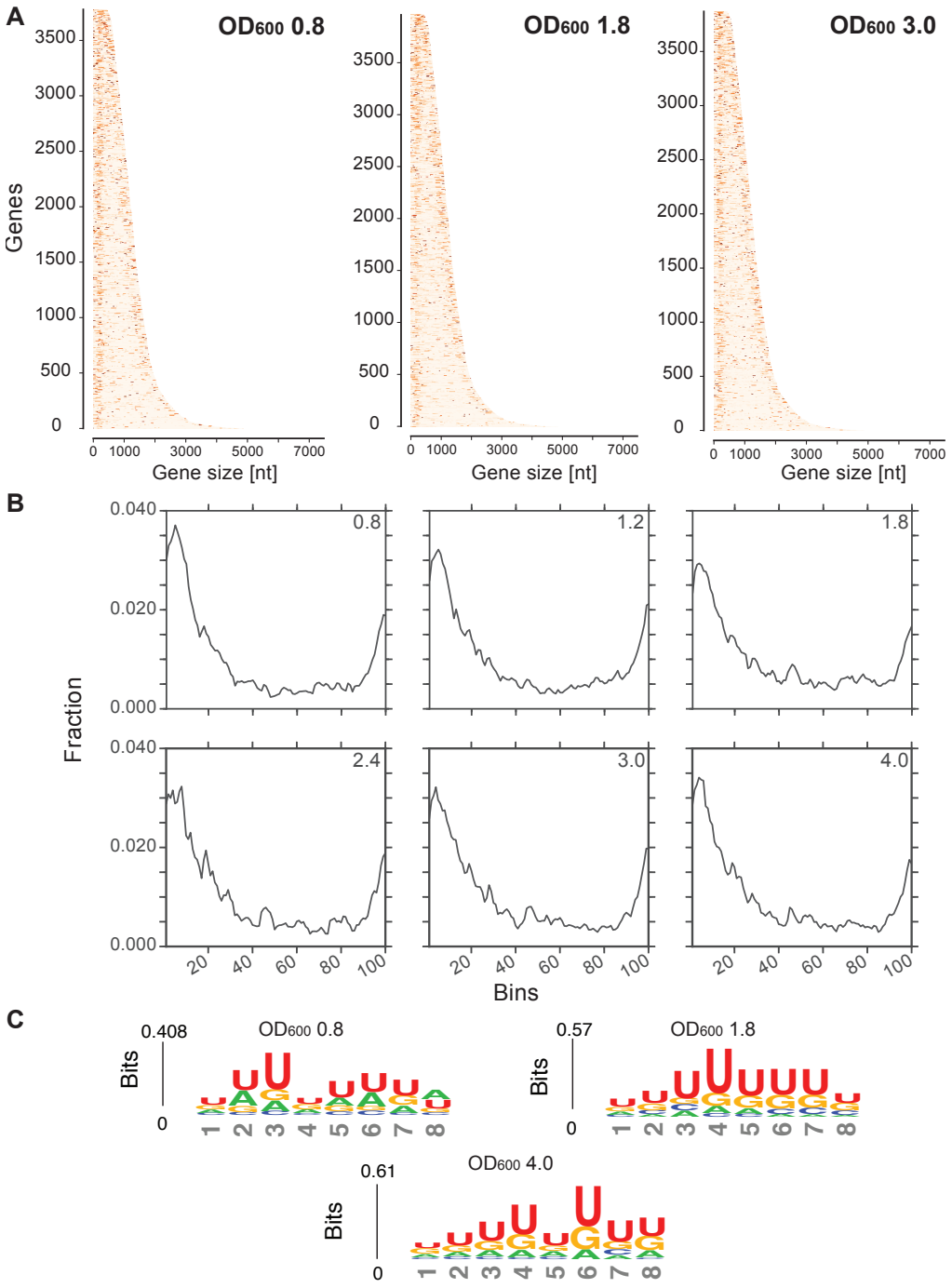

Fig. 3.12: Transcriptome-wide maps of Hfq binding to protein-coding genes. (A) Heatmaps showing the distribution of Hfq binding sites across all protein-coding genes at each OD_{600}: The genes are sorted by length (x-axis); the darker a nucleotide is, the more Hfq is bound to it; the cDNAs were flattened into clusters and a set UTR length of 200 bp was used for heatmap generation. (B) A more stringent selection of the genes used to generate the distribution of Hfq binding to the transcriptome: all genes with overlapping 5' or 3'-UTRs were removed from the analysis to avoid 'duplicate' counting; for all remaining cDNAs, FDR intervals of minimum 20 bases were considered for distribution plotting. The interval length (with UTR flanks as in the .gtf annotation file) for each gene was normalized over 100 bins as in Fig. 3.7. (C) Hfq binding motifs search in protein-coding genes: significant k-mers (4-8 nt in length) were identified using pyMotif tool of the pyCRAC package (Webb et al., 2014) and the motif logo was generated using all k-mers with a Z-score > 3, with kpLogo (Wu and Bartel., 2017).

3.8. Overview of the RNA-RNA interactions uncovered by CLASH

To analyse the RNA-RNA interactions captured by Hfq CLASH, I combined the data for all OD_{600} into one large dataset, comprising two CLASH experiments. The distribution of combinations of transcript classes found in all discovered chimeric reads indicates sRNA-mRNA interactions as the most frequent Hfq-mediated interaction type (~40%) (Fig. 3.13A). However, a large number of hybrid parts mapped to sRNA - intergenic regions and protein-coding -protein coding interactions.

To obtain an overview of the interactions recovered by Hfq CLASH compared to the RIL-seq data (Section 1.8), I flattened all chimeric reads into interactions, i.e an RNA-RNA pair was only counted once even if it was represented by many chimeric reads in any orientation (i.e RNA1-RNA2 and RNA2-RNA1 were counted as one interaction). At this moment, I do not have a completely working pipeline to select enriched interactions from the CLASH data, and adaptation of approaches used by other groups posed challenges. Since this part of the work is in progress, for the time being I set a threshold of minimum two unique hybrids to consider an interaction possible. The RIL-seq data was retrieved from the supplemental information provided with the publication (Melamed et al., 2016) and includes significant interactions for the comparable growth conditions tested (exponential phase of growth and stationary phase). The chimeras containing protein-coding genes in at least one part of the read were compared for Hfq CLASH and RIL-seq. For both datasets, due to annotation differences, intergenic regions and insertion elements were removed from the analyses.

The comparison shows that although performed in the same organism, and some growth conditions somewhat overlapped, only very few interactions are captured by both approaches, roughly a 13% ratio from all interactions of each dataset. Even without statistical filtering, the distribution profiles of chimera parts across protein-coding genes can be a good indication of how specific the sRNA-mRNA capture was using CLASH. To generate distribution profiles for chimera parts, clusters containing

only the chimera parts that uniquely mapped to single genes were considered and were counted only once. The genomic sequence was divided in 100 bins and the hit density in each bin was calculated as before (Fig. 3.12 B). The distribution across all protein-coding genes (Fig. 3.14 A, left) shows that most chimera parts map to the 5'-UTRs and some enrichment was also found in 3'-UTRs of mRNAs. This pattern is consistent with Hfq binding sites distribution (Fig. 3.12 A-B). Next, to check whether there are any chimeric reads parts overlapping the AUG codon region, I have similarly considered only the mRNA clusters that uniquely mapped to single genes, and counted them only once, but generated bins from the 5'-end of coding-sequence coordinates (Fig. 3.14 A, right). The resulting distribution shows that chimera parts are enriched in the AUG codon window, consistent with the canonical mode of translation inhibition by sRNAs.

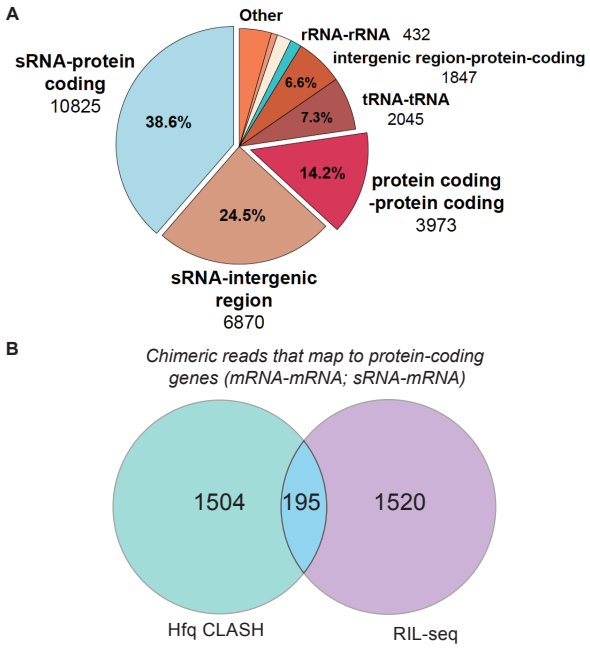

Fig. 3.13: Brief overview of the interactions captured by Hfq CLASH. (A) Intermolecular transcript combinations found in interactions captured by Hfq CLASH. Combination count of all uniquely annotated hybrids on genomic features. Note: a single interaction can be represented by many unique chimeras. (B) Comparison between Hfq CLASH data cumulative for all OD_{600} and RIL-seq data (Melamed et al., 2016). cumulative for all growth conditions: The Venn diagram shows the overlap between Hfq CLASH interaction pairs (minimum two chimeras per interaction) and RIL-seq significantly-enriched interactions. Each interaction is counted only once in each dataset, regardless of how many hybrids support the interaction. Due to annotation differences, intergenic regions and insertion elements were removed from the comparative analyses.

Next, to distinguish between the motifs enriched in 5'-UTR of protein coding-genes and 3'-UTR regions, I generated clusters uniquely mapping to these regions and searched for enriched K-mers (Fig. 3.14B). The significantly enriched K-mers point to distinct motifs characteristic to 5'-UTR and 3'-UTR regions. As for Hfq, the 3'-UTR-containing chimera consensus motif is enriched in polyU tails, whereas the A-rich 5'-UTR motif is more consistent with the (ARN)n motif proposed for Hfq by previous studies (Tree et al., 2014).

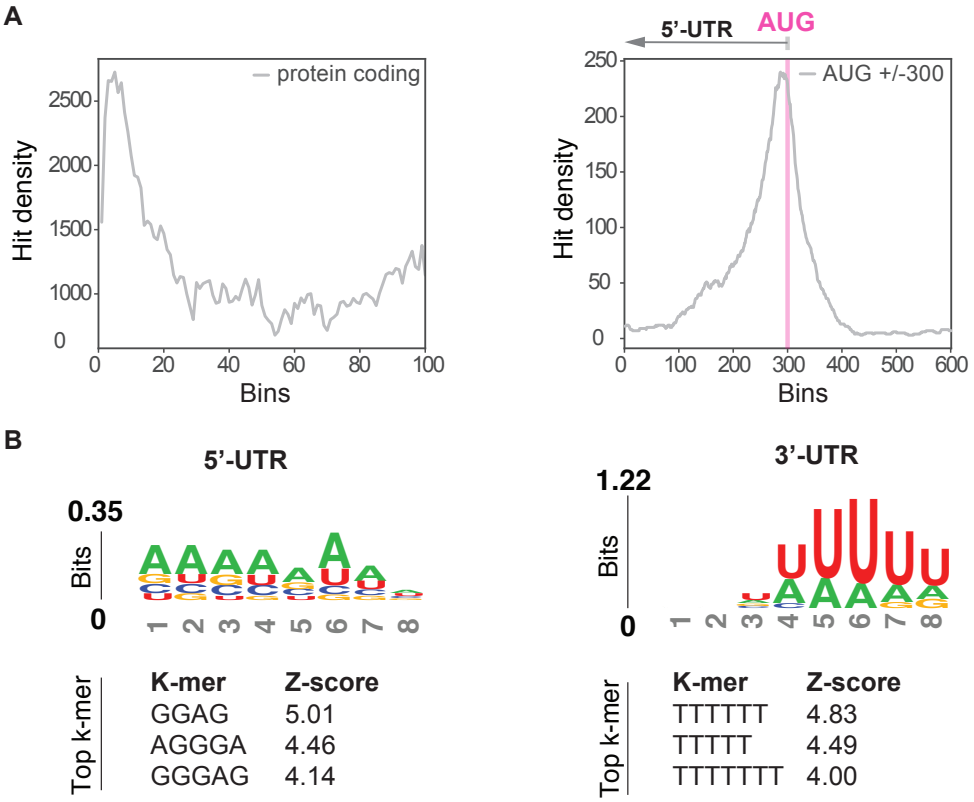

Fig. 3.14: Analysis of the clusters of chimeras with parts that map to protein-coding genes. (A) Distribution of clusters that uniquely map to protein-coding genes. The parts of the chimeras were collapsed into clusters; distribution plots were generated by removing the clusters that map to multiple genes and (using bedtools), and the remaining clusters were counted only once. The distribution plot over protein-coding genes was generated by dividing each gene in 100 bins and calculating the hit density (y-axis); For the distribution plot around the AUG, the gene length was normalized in 601 bins (x-axis) 5'-end overlap (-300) before the start of the coding sequence, and 300 bins downstream AUG (+300); the bins corresponding to the start codon are indicated in pink. (B) Enriched motifs in chimeras that uniquely overlap the 5'-UTRs and 3'-UTR; the logo was made as in Fig. 3.12 using K-mers with a Z-score > 2. The top K-mers after the pyMotif tool search are shown below each logo.

3.9. Hfq CLASH rediscovers known sRNA interactions and uncovers novel partners for model sRNAs

To test the robustness of the Hfq CLASH methodology in uncovering true intermolecular RNA interactions, I compared the Hfq CLASH data with sets of known, experimentally validated sRNA-mRNA pairs.

I performed a detailed analysis of the hybrid reads recovered in the Hfq CLASH data with the model sRNA ChiX. ChiX is an abundant sRNA at all stages of growth (Fig. 3.2) that was shown to downregulate the expression of *chiP* by binding to its SD sequence to prevent translation initiation (Fig. 3.15A) (Bossi et al., 2009; Overgaard, 2009; Plumbridge et al., 2014). The parts of ChiX-*chiP* chimeras in the CLASH data that mapped to *chiP* support the described mechanism of repression - all chimera parts that support the interaction mapped to the correct genomic location (5'-UTR and start codon region) of *chiP*. Moreover, the hybrid structure predicted by RNACofold using the sequences found within the chimeras (Fig. 3.15C), further confirmed that the base-pairing of ChiX to the SD region of *chiP* is responsible for repression.

I have also objectively re-discovered in the CLASH data the interaction between ChiX and *chbBC*, an mRNA encoded within the *chbBCARFG* operon. *chBC* RNA can act as both a decoy and target of ChiX regulation (Plumbridge et al., 2014). In the absence of chitosugars, ChiX was shown to downregulate the expression of *chbC* by binding to the intercistronic region between *chbB* and *chbC*. Consistent with this mode of regulation, the CLASH hybrid reads contained ChiX chimeras whose cognate halves mapped to *chbBC* at the exact proposed location (Plumbridge et al., 2014). Additionally, I found in the Hfq CLASH data some novel ChiX targets. The CLASH data contains ~900 hybrid reads with parts that mapped to ChiX (Fig. 3.15 B, top panel; Fig. 3.16 A). The interactions supported by at least 5 unique chimeras place ChiX in a network that includes a roughly equal number of ChiX-mRNA and ChiX-sRNA interactions (Fig. 3.16A). I have dissected two examples of these hybrids

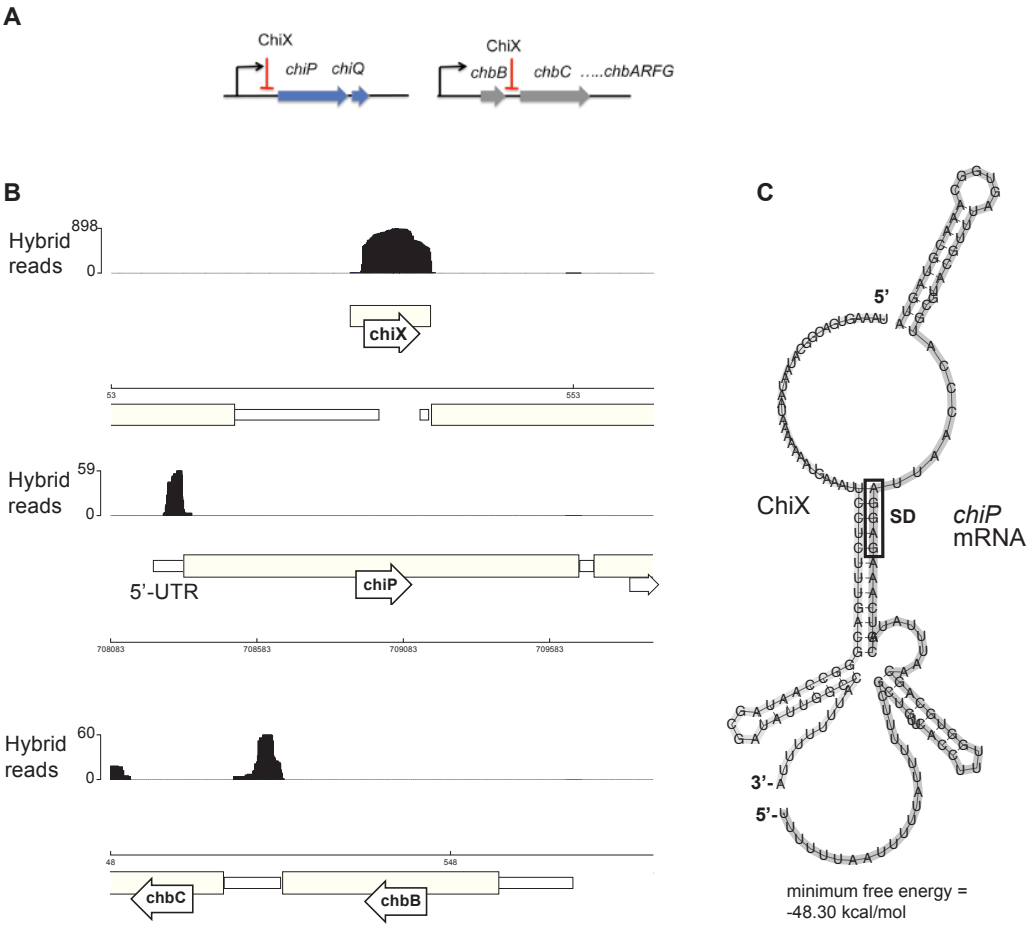

Fig. 3.15: Example of re-discovery of known sRNA-mRNA interactions by Hfq CLASH. (A) Diagram of known ChiX sRNA-mediated regulation. ChiX is an abundant sRNA that down-regulates *chiP* expression by occluding the Shine-Dalgarno (SD) sequence. In the absence of chitosugars, ChiX also represses *chbC* by base-pairing to a intercistronic region between *chbB* and *chbC*. Image adapted from Plumbridge et al., 2014. (B) Genome-browser plot showing the location and peaks of hybrid reads for ChiX (the counts on the y-axis correspond to all ChiX-containing hybrids recovered by Hfq CLASH); below plots for two of its known targets, *chiP* and *chbC* are shown (the hybrid read count are specific to ChiX-*chiP* and ChiX-*chbC* hybrids, respectively. (C) Prediction of the base-pairing between ChiX and *chiP*, using the RNA sequences recovered by Hfq CLASH and RNACofold (Lorenz et al., 2011). The SD sequence within the *chiP* mRNA is highlighted.

to verify if these are de *bona fide* interactions. The analysis of the hybrids of ChiX with a new mRNA target, *mreC*, shows that the genomic mapping of the ChiX-*mreC* chimeras overlaps the 5'-UTR and first few codons of *mreC* mRNA (Fig. 3.16 B). At a first glance, this points to a canonical mode of translational repression executed by ChiX. This regulatory mode is further supported by the favourable duplex structure predicted *in silico* (Fig. 3.16 C), which suggests that ChiX base-pairing to *mreC* may block access to the AUG start codon and a few nucleotides downstream, an interference that typically results in inhibition of translation (Section 1.4).

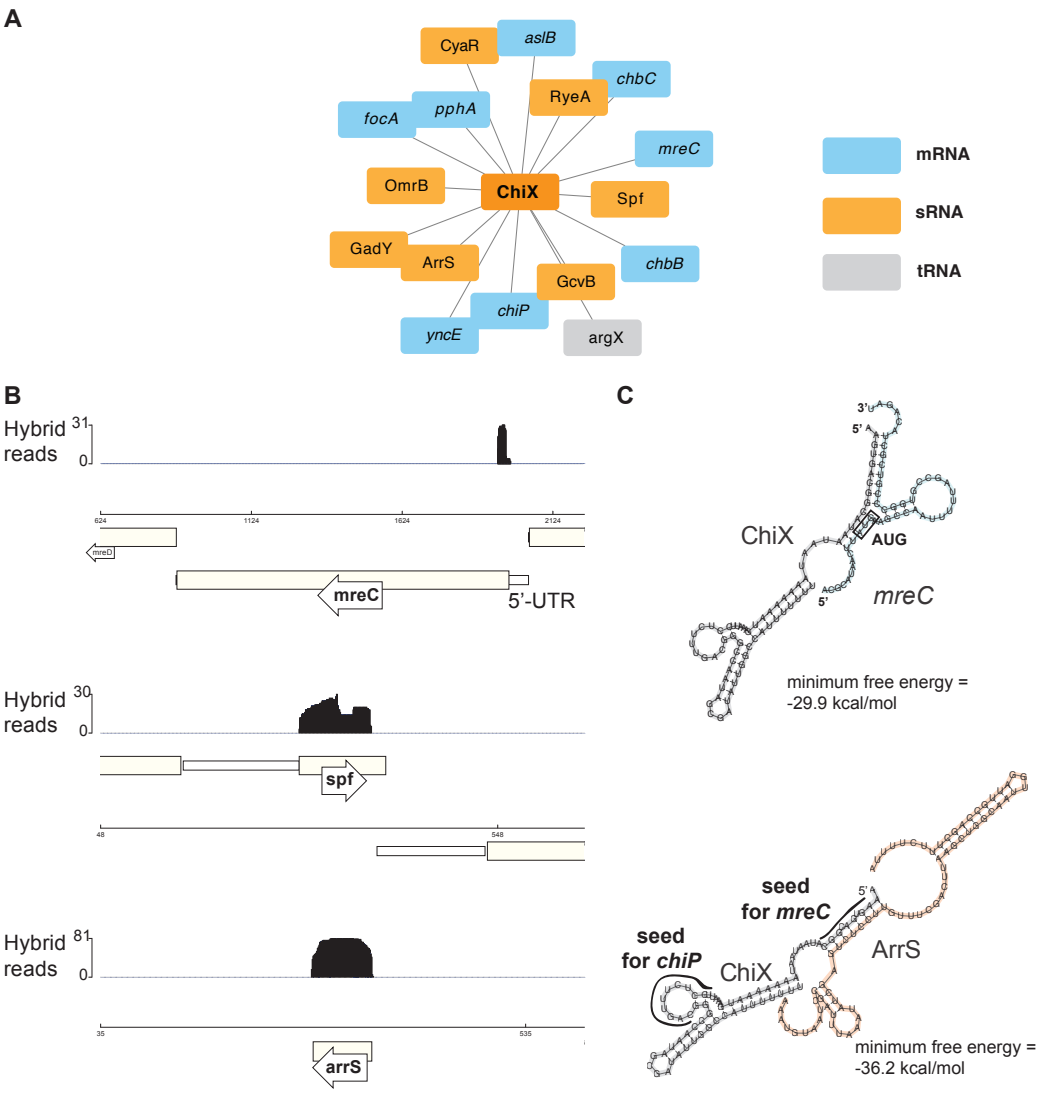

Fig. 3.16: Example of discovery of new targets of well-characterized sRNAs by CLASH. (A) A network showing direct interactions between ChiX sRNA and other RNAs, supported by at least 5 hybrids in the Hfq CLASH data; drawn with Cytoscape (Shannon et al.,2003). (B) Genome browser plots that illustrate the location of the parts of ChiX-containing chimeras that map to *mreC*, *spf* and *arrS*, respectively. Hybrid counts for each ChiX-RNA interaction are labelled on the y-axes. (C) Hybrid structures of ChiX base-pairing with targets uncovered by CLASH, predicted using RNAcofold (Lorenz et al., 2011). In the ChiX-*mreC* hybrid structure, the start codon of *mreC* is highlighted (AUG). In the ChiX-ArrS hybrid structure, the seed sequence utilized or predicted to be utilized (B) by ChiX to base-pair with *chiP* and *mreC* mRNAs are highlighted with black lines.

Intriguingly, ChiX also appears to base-pair with sRNAs. Abundant hybrids were found in support of ChiX-Spf and ChiX-ArrS interaction (Fig. 3.16 B). I analysed the ChiX-ArrS interaction in more detail, because ArrS, known as a *cis*-encoded sRNA was found to bind Hfq (Fig. 3.9). The interaction is supported by a very high number of unique chimeric reads (81). Interestingly, the predicted duplex structure between the two sRNAs masks the ChiX-binding site predicted to base-pair with *mreC*, whereas

the ChiX seed for *chiP* interaction remains available (Fig. 3.16C). This suggests that the ArrS sRNA could specifically pair with ChiX to de-repress *mreC* expression.

3.10. sRNA-sRNA interactions are abundant in *E. coli*

The capture of multiple ChiX-sRNA interactions by Hfq CLASH motivated me to search for more sRNA-sRNA hybrid reads in the CLASH data. First, I selected the sRNA-sRNA interactions supported by at least two hybrid reads in the Hfq CLASH data and intersected the sets of interactions with available RNase E CLASH (Waters et al., 2017) and RIL-seq (Melamed et al., 2016) datasets (Fig. 3.17). I retrieved these data from the respective publications, with no additional filtering. However, for the analyses, I only considered the validated sRNAs that are transcribed from independent promoters, and are expressed in all *E.coli* species considered. I excluded 3'-UTR-derived sRNAs from the analysis because it was not possible to discriminate whether an sRNA-3'UTR duplex reflects sponging activity or sRNA-mediated regulation at 3'-UTRs mRNAs. I speculate that such duplex formation may also be a pathway for releasing sRNAs from the parent mRNAs. By considering strictly the sRNAs that do not overlap mRNAs, I could filter for sponge-sRNA interactions. Therefore, I specifically selected these sRNAs from all datasets to make the comparison. The biggest overlap was between sRNA-sRNA pairs identified in the Hfq CLASH data and RNase E CLASH data performed in EHEC *E. coli.* These sRNA-sRNA interactions are most likely true interactions (especially since the RNase E CLASH data was statistically filtered; Waters et al., 2017), but given the differences in growth conditions (virulence inducing conditions for EHEC *E.coli* and growth transitions for *E.coli* MG1655), these interactions may have roles in a wide range of physiological conditions. However, the pools of sRNA-sRNA interactions recovered in association with RNase E are expected to be different, because RNase E associated pairs only constitute a subset of all sRNA-sRNA interactions that specifically result in degradation of (at least one of) the sRNAs. The sRNA-sRNA interactions captured with Hfq, but not with RNase E are most likely growth-condition specific interactions.

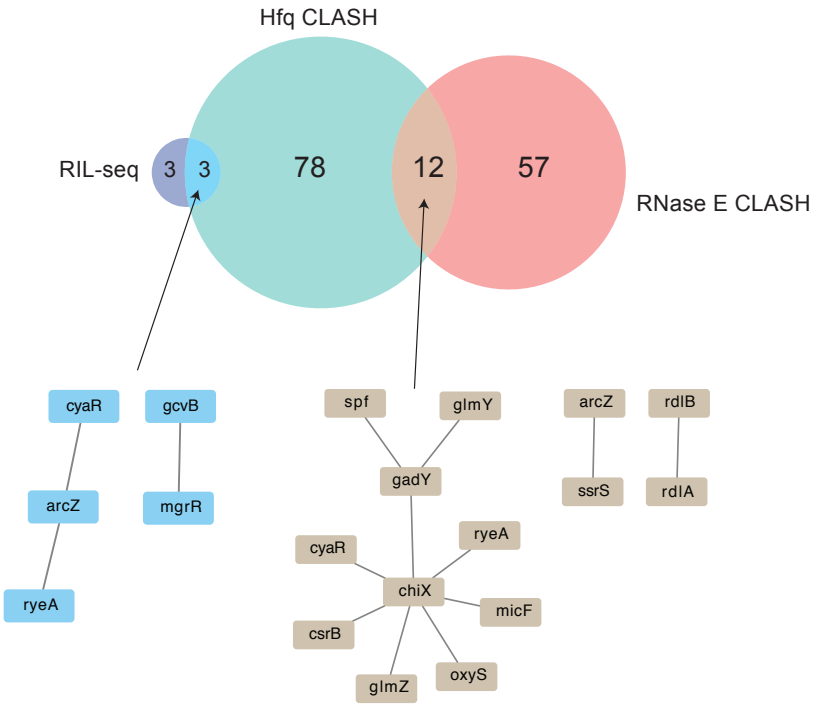

Fig. 3.17: Comparison of three datasets that mapped RNA-RNA interactions in *E.coli*: sRNA-sRNA interactions recovery. The Venn diagram shows the intersection of sets of sRNA-sRNA interactions independently identified by RIL-seq, RNase E CLASH and Hfq CLASH. The search for sRNA-sRNA hybrids was performed for known sRNAs, and did not include sRNAs derived/processed from parental mRNAs. The RIL-seq and RNase E CLASH interactions were retrieved from the tables of significant interactions provided with the publications (Melamed et al., 2016; Waters et al., 2017) without any additional filtering. The Hfq CLASH sRNA-sRNA interactions considered were those with a hybrid count of minimum two.

Interestingly, among the sRNA-sRNA interactions that associate with both Hfq and RNase E is a ChiX-CsrB interaction (Fig. 3.17), which suggests that one of the two sRNAs may direct the other to be degraded by RNase E, or only traps it in a complex.

A very low number of sRNA-sRNA pairs were uncovered by both Hfq CLASH and RIL-seq (3), even though both methods used Hfq as a bait for RNA-RNA isolation and some of the growth conditions were, in theory, comparable (i.e. exponential growth and stationary phase). However, cross-validation renders these interactions as highly likely to be occurring *in vivo*. Among these interactions is the GvcB-MgrR interaction, which links lipopolysaccharide composition to amino acid metabolism.

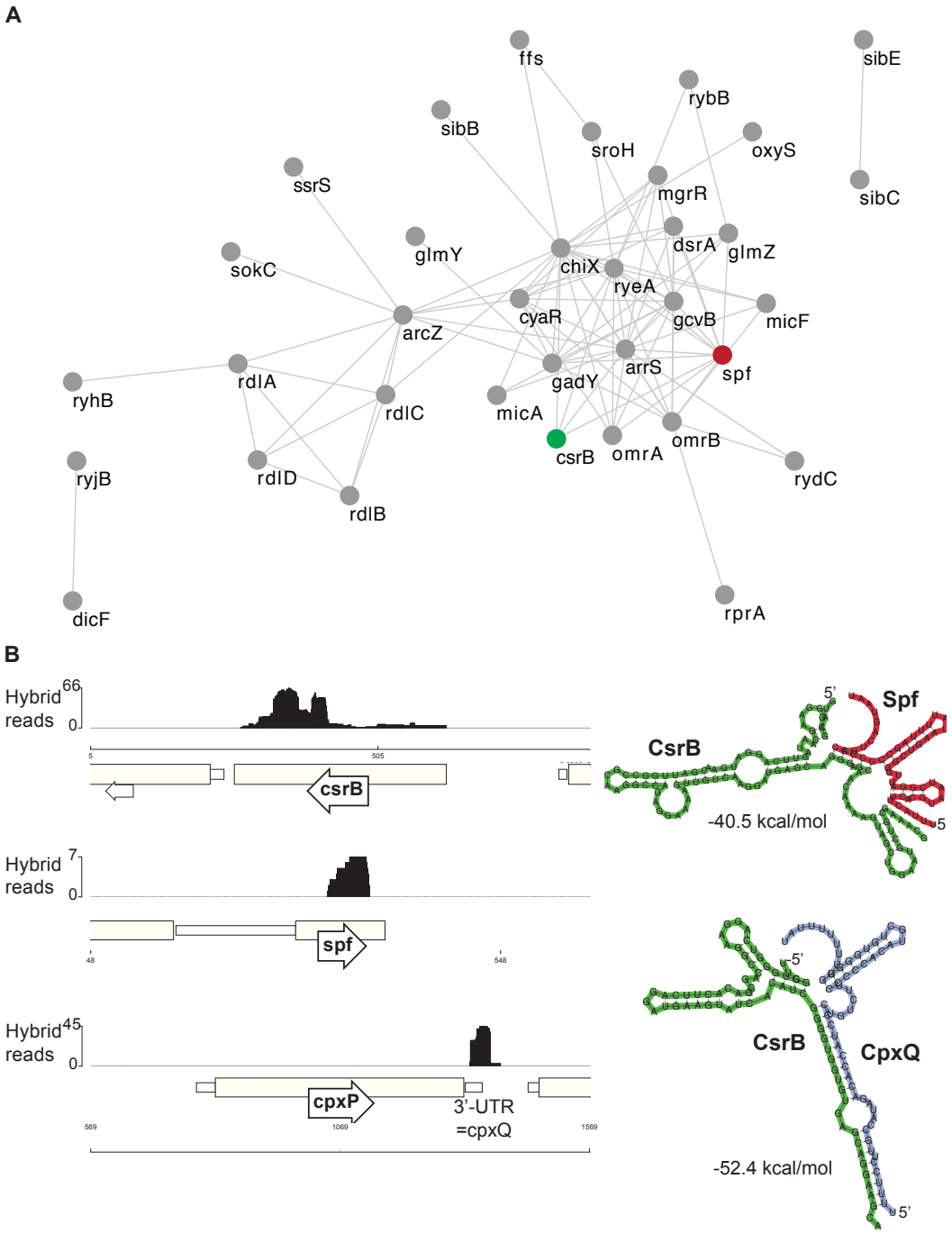

Fig. 3.18: sRNA-sRNA interactions captured by Hfq CLASH. (A) Network of sRNA-sRNA interactions captured by Hfq CLASH (drawn using Cytoscape; Shannon et al., 2003). sRNA-sRNA interaction sets were obtained as described in Fig. 3.17. The nodes in green (CsrB) and red (Spf) are highlighted because they are discussed in more detail. (B) Genome-browser plots that show the mapping location and hybrid counts of the peaks corresponding to CsrB-Spf and CsrB-*cpxP* 3'-UTR hybrids. The CsrB hybrid count includes the other CsrB-containing chimeras. On the right, the corresponding *in silico* predicted structure (RNACofold; Lorenz et al., 2011) for each interaction is shown, together with the free energy of binding (kcal/mol)

Next, I analysed in more detail the sRNA-sRNA interactions within the Hfq CLASH data (Fig. 3.18 A). The network of all sRNA-sRNA interactions represented by at least two unique chimeras illustrates that some sRNAs are hubs in this sRNA dominated circuitry: ChiX, ArrS and GcvB are such sRNAs, and this correlates with the fact that they are some of the strongest Hfq-crosslinked sRNAs (Fig. 3.3).

Since CsrB has not been previously described as an Hfq-dependent sRNA, I focused more on its association with other RNAs (Fig. 3.18 B). First, I analysed the chimeric reads that support CsrB-Spf interactions. Although the interaction (supported by 7 unique hybrids) is predicted to encompass only 5 base-pairs, all of them are contiguous GC pairs, which renders it a stable duplex structure. The GGA containing loops within CsrB are still exposed, suggesting it is more likely that CsrB binding to Spf may not be directly involved in CsrA binding to CsrB, but rather, CsrB sponges Spf.

Next, I analysed the interaction between CsrB and CpxQ, a recently discovered 3'-UTR derived sRNA (Chao and Vogel, 2016). The complementarity region as indicated by the predicted structure is unusually extended and GC-rich and would readily result in inactivation of CpxQ. For both Spf and CpxQ, however, it must be experimentally established that CsrB acts as the sponge.

3.11. 3'-UTRs of mRNAs are reservoirs of small RNAs

The enrichment of Hfq binding to the 3'-UTRs of genes and the identification of a poly-U motif in mRNAs within chimeras with parts that map to the 3'-UTRs, suggests that Rho-independent terminator-containing mRNA regions are extensively involved in Hfq-mediated RNA-RNA interactions. Since a few isolated studies characterised discrete sRNAs that overlap the 3'-UTRs of mRNAs (Guo et al., 2014; Myiakoshi et al., 2015a), and global studies generated lists of potential mRNAs that may contain sRNAs in their 3'-UTRs (Chao et al., 2017), it has become increasingly recognized that sRNA biogenesis from 3'-UTRs is a widespread phenomenon (Section 1.7).

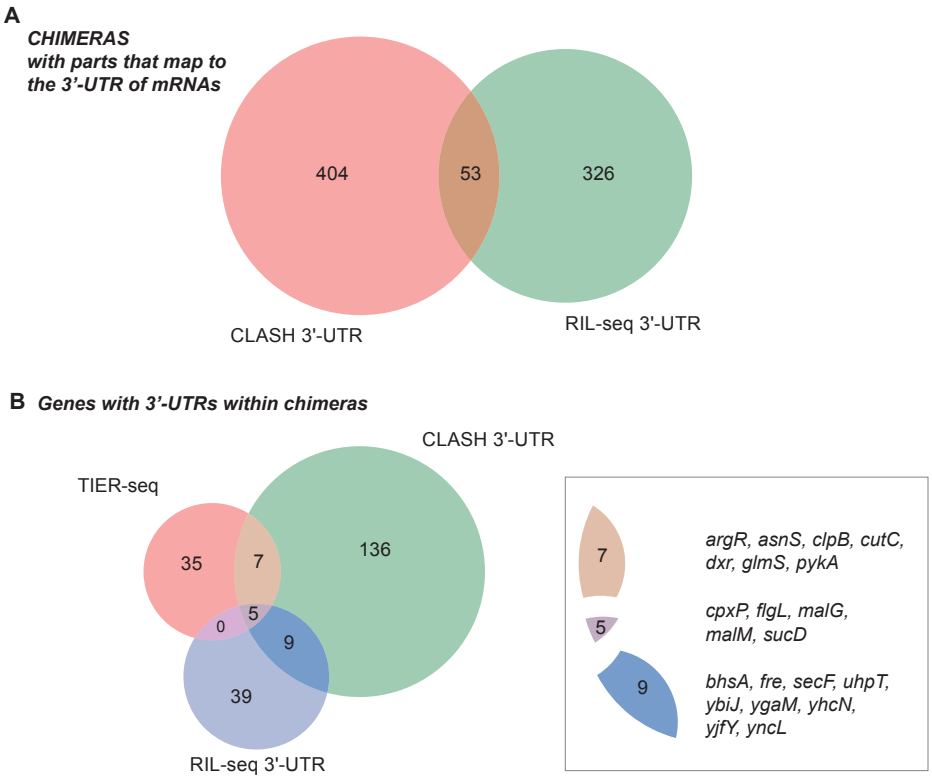

Fig. 3.19: 3'-UTR containing chimeras may uncover novel 3'-UTR-derived sRNAs. (A) Venn diagram comparing the sets of interactions with one part of the chimera mapping to a 3'-UTR in the Hfq CLASH and RIL-seq datasets. The RIL-seq RNA-RNA interaction set was obtained as described in Fig. 3.17, and filtered for the 3'-UTR annotation on either orientation of the RNA-RNA pairs. Differences in annotation between the two datasets were not accounted for, e.g.. if in the CLASH dataset a certain region is annotated as 3'-UTR, but in the RIL-seq data it is annotated as an intergenic region (IGR or IGT), this region will be underrepresented in the RIL-seq data used for set intersection and Venn diagram plotting. (B) The names of genes that are part of chimeras, and also uniquely map to 3'-UTRs, were selected from the CLASH data and RIL-seq data. Both were intersected with the set of mRNAs that were predicted by TIER-seq studies (Chao et al., 2017) to harbour sRNAs in their 3'-UTRs, that get released by RNase E processing.

Given that 5'-UTRs are usually more often targeted during sRNA-mediated regulation, I speculate that many hybrid reads containing 3'-UTRs of mRNAs, especially those within mRNA-mRNA chimeric reads, actually involve a 3'-UTR derived sRNA. To further investigate this hypothesis, I selected all hybrids in the Hfq CLASH data that contain at least one part of the chimeric read that mapped to 3'-UTRs. I then intersected this set of 450 chimeras with the set of chimeras containing 3'-UTR fragments from the RIL-seq data (Fig. 3.19 A). As for all comparative analyses, I only considered the interactions with at least two hybrid counts in the CLASH data. Although many 3'-UTR-containing hybrid reads are not shared by both datasets, 53

pairs are identical, indicating that they may be bona fide interactions - and that at least some of them may contain 3'-UTR-derived sRNAs.

To further investigate which mRNA 3'-UTRs are more likely to be sRNAs, I extracted the gene names of all mRNAs whose 3'-UTRs were involved in hybrids - this also includes 3'UTR-3'UTR chimeric reads, so for these interactions it is hard to establish which 3'-UTR is the regulator. However, more clear cases in which 3'-UTR derived fragments form chimeras with the 5'-UTR of other mRNAs can streamline regulator identification.

One approach that can help restrict the lists of 3'-UTR derived candidate sRNAs, besides cross-validation with the RIL-seq data, is the intersection with the sets of mRNAs predicted to contain sRNAs that are released by RNase E processing, uncovered by TIER-seq in *Salmonella* (Chao et al., 2017) (Fig. 3.19 B). Although performed in a different bacterial species, *E. coli* and *Salmonella* are related and share many post-transcriptional regulators, so the insights gained from comparing these datasets may well apply to both organisms. The results of the analysis comparing the CLASH data, RIL-seq data and TIER-seq data revealed that only 5 mRNA 3'-UTRs are uncovered in all datasets - however these are very likely candidates for 3'-UTR -release as sRNAs. The Hfq CLASH and RIL-seq datasets may include 3'-UTR-overlapping sRNAs that either are transcribed from independent promoters, and sRNAs that may require a different RNase to be released from the parental mRNAs.

I analysed in more detail two 3'-UTR-overlapping candidate sRNAs that were uncovered in both datasets. One of them is the *cpxP* 3'-UTR (Fig. 3.20 A). Indeed, the 3'-UTR of this mRNAs was shown to accumulate as a sRNA upon RNase E-mediated processing and this stable fragment acts as a functional sRNA (Chao and Vogel., 2016). Although this sRNA was identified in both CLASH and RIL-seq datasets in a large number of chimeric reads, most of the CpxQ targets captured by each method do not overlap, suggesting again that these two datasets are complementary. Either method greatly increased the set of targets for this sRNA, and shown are two examples that link this sRNA to control of expression

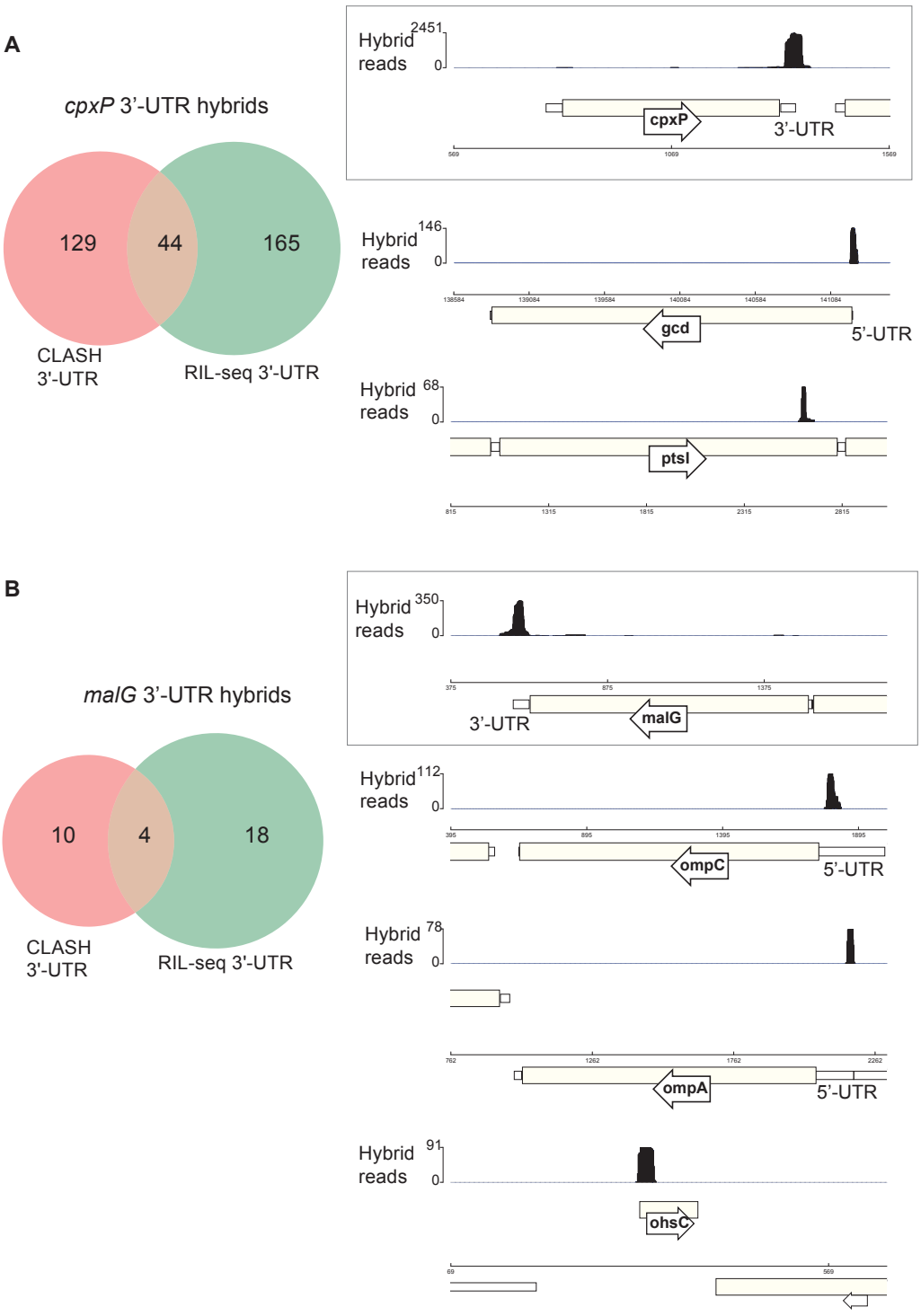

Fig. 3.20: 3'-UTRs of mRNAs that are strong sRNA candidates. (A) *cpxP*: Venn diagram showing the overlap between *cpxP*-containing chimeras that map to the 3'-UTR in RIL-seq and Hfq CLASH data, and genome browser plots showing the regions that *cpxP*-containing chimera parts map to, as indicated by the CLASH data. The plots surrounded by a box highlight the 3'-UTR-containing mRNA that is likely a source of 3'-UTR derived sRNAs, and the counts are cumulative of all hybrid reads that mRNA 3'-UTR is part of. (B) As in (A), for *malG* 3'-UTR as a candidate source of sRNAs.

of components of carbohydrate transport systems (*ptsI*) and glucose oxidation (*gcd)*. CpxQ appears to employ two different mechanisms for regulating these targets, as it binds to *gcd* mRNA at a site in the 5'-UTR, and to *ptsI* mRNA in the coding sequence, but both base-pairing events may be followed by RNase E processing promoted by the 5'-P end of CpxQ.

An uncharacterized 3'-UTR that may be released by RNase E processing originates from *malG* (Fig. 3.20 B). As for CpxQ, the targets do not completely overlap between the CLASH and RIL-seq datasets. The *malG* 3'-UTR-containing chimeras were most abundant with two mRNAs encoding major porins, *ompC* and *ompA,* to which *malG* base-pairs deep within the 5'-UTR, possibly causing their degradation promoted by the processed 5'-end of this sRNA. This 3'-UTR was also found in chimeras with a *cis*-encoded sRNA, OhsC, suggesting that either *malG* controls its expression by sponging, or vice-versa. The most abundant and favourable interactions (with *ompC* and *ompA*) appear to be utilizing roughly the same region of MdoR for base-pairing (Fig.3.20 B), indicative that the corresponding site on the sRNA may be a main, functional seed. Similarly, the targeted mRNAs base-pairing regions mapped to the 5'-UTRs (*ompC, ompA)*. Thus, these data indicate that MdoR is a classical trans-acting in *E.coli* that could mediate RNA decay by recruiting Hfq and RNase E (Pfeiffer et al., 2009), cause translational inhibition (Bouvier et al., 2008) or affect premature terminate transcription of these mRNAs (Sedlyarova, 2016).

Finally, I analysed a 3'-UTR derived candidate sRNA that was captured by Hfq CLASH only (Fig. 3.21). The 3'-UTR of *ahpF*, coding for a hydroperoxide reductase shown to be also involved in resistance to aminoglycoside antibiotics (Ling et al., 2012). Its 3'-UTR was found in chimeric reads with a few other mRNAs (Fig. 3.21 A), among them *recC* (involved in stress-induced mutagenesis; Al et al., 2012) and *aroE*. For both, the chimeric reads map to the coding sequences, although for *aroE* (involved in aromatic amino acid biosynthesis) the binding region extends into its 3'-UTR and encompasses a long region of complementarity.

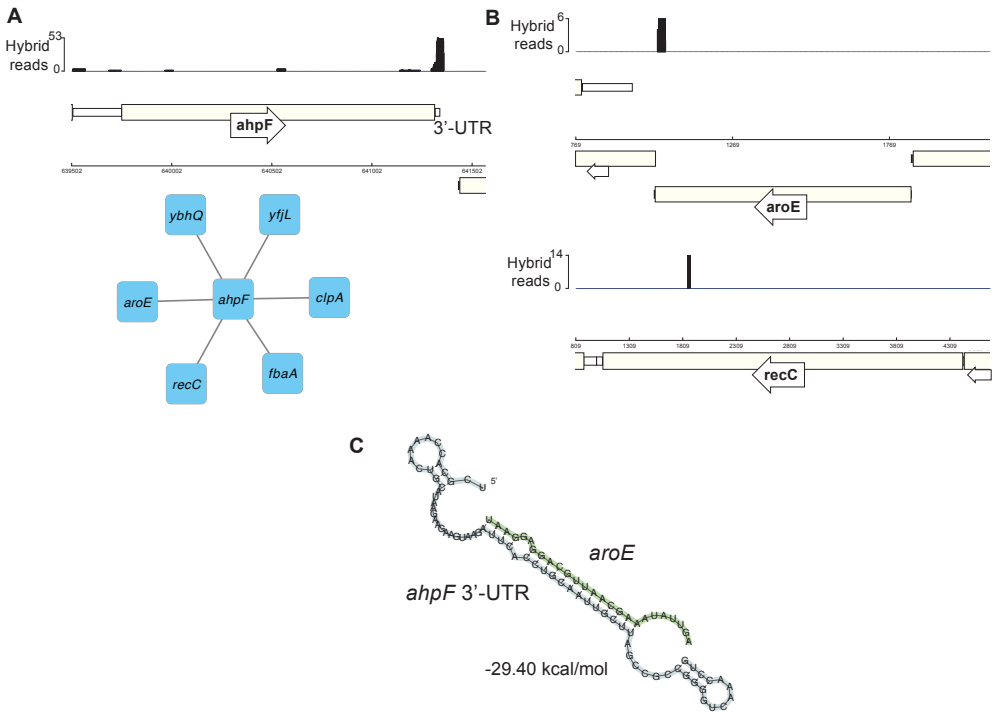

Fig. 3.21: An example of a candidate 3'-UTR-derived sRNA that was not captured by either RIL-seq or TIER-seq. (A) Genome-browser snapshot showing the total number of hybrids the *ahpF* 3'-UTR is part of (y-axis). Below, a network with its prime interacting mRNAs is shown (minimum 5 hybrids) (drawn with Cytoscape; Shannon et al., 2003). (B) Genome-browser image showing the number of *ahpF-recC* and *ahpF-aroE* hybrids and where the corresponding parts of the hybrids map to within these mRNAs. (C) *In silico* predicted structure of *ahpF* 3'-UTR with *aroE* 3'-UTR/intercistronic region, using RNACofold (Lorenz et al., 2011)

3.12. Validation of novel sRNAs by Northern blot

Following the analysis of Rho-independent terminator-containing intergenic regions found closely downstream the coding sequence of some mRNAs, and some of the hybrid reads containing 3'-UTRs of coding genes, I validated a few candidate sRNAs by Northern Blot (Fig. 3.22). Both the OD_{600}-dependent Hfq binding and RNA steady-state profiles revealed that the dynamics of expression are defined throughout growth and can vary significantly from one sRNA to another. For instance, INT-0-1913 shows a gradual increase in expression with OD_{600}, whereas INT-0-2492, now referred to as GadF (Melamed et al., 2016), is highly expressed mainly at high OD_{600} (Fig. 3.22). The dynamic expression profiles of these sRNAs indicated that they may have a role in regulating adaptation to various growth stages. Strikingly, one of the candidate

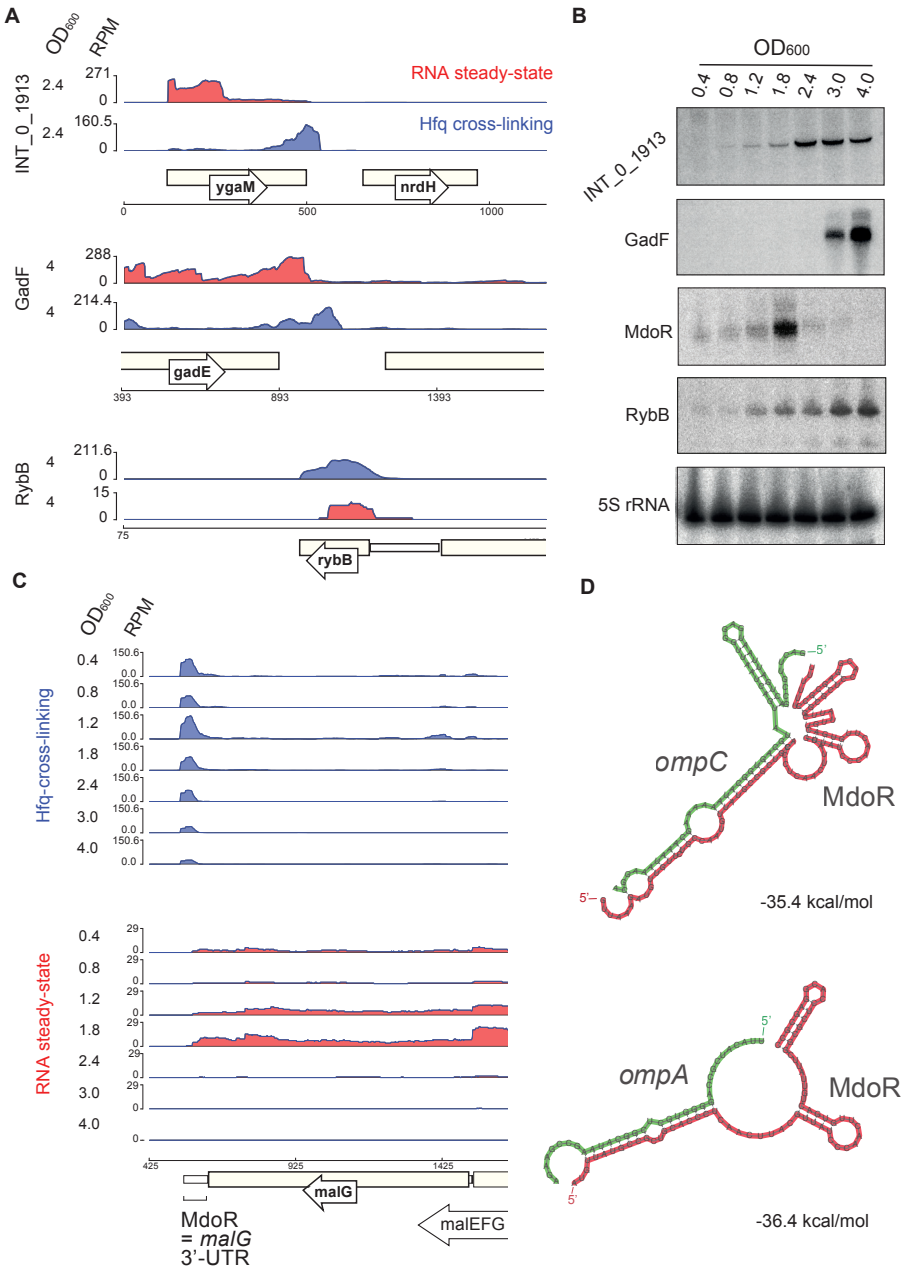

Fig. 3.22: A survey of the validated novel sRNAs uncovered by CLASH in *E.coli.* (A) Genome-browser plot of several regions containing candidate sRNAs for OD_{600} at which the RNA steady-state was maximal for each candidate; the candidate names and OD_{600} are indicated at the left side of each track-pair; the y-axis shows the normalized reads (RPM: reads per million); red: RNA steady states from the RNA-seq experiment, blue: Hfq cross-linking from the CLASH experiment; (B) Northern blot validation of candidate sRNAs;, RybB was probed as a characterized sRNA positive control and 5S rRNA as the loading control.(C) Genome-browser plots showing the Hfq crosslinking (CLASH experiment; blue) and RNA steady-states (RNA-seq experiment; red) for MdoR (*malG* 3'-UTR) at different growth-stages; the y-axis indicates normalized reads (RPM: reads per million); (D) *In silico* prediction (RNACofold, Lorenz et al., 2011) of hybrid structures derived from the most abundant MdoR chimeric reads with mRNAs

sRNAs, here referred to as MdoR (*mal*-dependent OMP repressor) is very transiently expressed at OD_{600} 1.8, i.e., at the transition from the exponential phase to the stationary phase of growth. This is one of the main reasons why I chose to study the function of MdoR in greater detail (described in Chapter 4).

In silico predictions of the MdoR-*ompC*, MdoR-*ompA* duplex structures (RNACofold; Lorenz et al., 2011) revealed that MdoR can form extensive and highly energetically favourable base-pairing interactions with these targets (Fig. 3.22 D). The most striking hybrid structure was the one between MdoR and *ompC* 5'-UTR, encompassing a region of 24 bp - such extensive duplex structures are not typically found in *trans*-encoded sRNA regulation in *E.coli*. Consistent with my CLASH data, the RIL-seq approach (Melamed et. al, 2016) also uncovered hybrids between *ompC* and the 3'-UTR of *malG*.

3.13. Discussion

The analyses performed in this Chapter uncover many exciting prospects of sRNA-mediated regulation and sRNA networks in bacteria. Many of the insights ensue from the fact that we profiled these interactions in multiple conditions of growth, rather than in a single condition.

Transcript classes bound by Hfq and involved in RNA-RNA interactions

I have found that all transcript classes can associate with Hfq and form intermolecular chimeras (Fig. 3.2; Fig. 3.13 A).

The distribution of cDNA reads among transcript classes (Section 3.2) was a first indication that Hfq binding to RNAs is correlated with growth. The pattern of distribution was similar to the distribution of transcript classes uncovered by Hfq CRAC in enterohemorrhagic *E. coli* (Tree et al., 2014). The association of sRNAs and mRNAs is present at all optical densities tested, indicating that all growth

phases require sRNA-mediated regulation. Consistent with this, the largest fraction of hybrid reads recovered in the Hfq CLASH data were sRNA-mRNA interactions (Fig. 3.13 A). An increase in Hfq cross-linking to sRNAs and mRNAs with OD_{600} can be partially explained by the increase in sRNA-mediated regulation associated with stress responses triggered as cells transition from exponential stage of growth to the stationary phase of growth (Rau et al., 2015).

Some sRNAs are known to be induced at the stationary phase of growth or to regulate the expression of RpoS, the general stress (stationary phase) sigma factor (Gottesman and Storz, 2011). Interestingly, the growth stages dominated by Hfq binding to sRNAs (OD_{600} 1.8-2.4) was at the transition between the exponential and stationary phases of growth. The sRNAs occupying Hfq at these OD_{600} may have specific expression patterns and may play important regulatory roles during growth, e.g shifting metabolic strategies, membrane composition changes etc. Other transcript classes were also bound by Hfq, most notably tRNAs and rRNAs - this observation is not surprising, given the proposed roles for Hfq in tRNA modification (Lee and Feig., 2008) and the discovery of tRNA sponges (Lalaouna et al.,2015). We also uncovered tRNA-tRNA hybrids (~ 7% of all hybrids) and rRNA-rRNA hybrids. However, these transcripts show a decline in association with Hfq as the cell density increases. The concomitant increase of Hfq binding to sRNAs and mRNAs, corroborated with the fact that the Hfq protein pool is limited (Moon and Gottesman, 2011), suggests that sRNA-mRNA pairs binding to Hfq may be prioritized over Hfq association with rRNA and tRNAs as the cells start reaching stationary phase. I speculate that the decrease in Hfq binding to rRNA and tRNA may render more Hfq hexamers available for sRNAs and mRNAs or sRNAs and mRNAs out-compete these transcripts for Hfq binding.

The growth-dependent patterns of Hfq association with sRNAs may reveal atypical roles for sRNAs

The analyses I performed to obtain growth stage-resolved snapshots of Hfq-association with sRNAs (Section 3.3, 3.4 and 3.6) and how well it correlates with

changes in total RNA levels revealed that globally, these two are not always positively correlated. The positive correlation improves as the cells approach stationary phase (from OD_{600} 2.4). Individual sRNAs that deviate from linearity are ArrS and AgrA (Fig. 3.6). Intriguingly, these are known as *cis*-encoded sRNAs (Aiso et al., 2014; Weel-Sneve et al., 2013), that should not require Hfq for their function. Therefore, the poor correlation of association of Hfq to these transcripts may occur because they perform Hfq-independent functions. However, I observed in my CLASH data a large number of chimeras with ArrS, many of them reflecting ArrS-sRNA interactions (Fig. 3.16 C, Fig. 3.18 A). One of these was with ChiX (Fig. 3.15). Corroborated, these data strongly suggest that ArrS can act as a sponge sRNA. Since under the tested growth conditions ArrS expression was induced at higher OD_{600}, possibly connected to the decrease in intracellular pH (Aiso et al., 2014), these interactions may help coordinate multiple pathways with the acid-response.

Two other interesting sRNAs are CrsB and CsrC. The Hfq-binding changes for these transcripts follow a slightly different profile than that of the change in the RNA levels. However, simply finding them associated with Hfq is striking, because these sRNAs are known exclusively for their role in sequestering the CsrA protein (Romeo et al., 1998, Liu et al., 1997). Strikingly, we found that Hfq binds roughly one third of the CsrA binding sites on CsrB (Fig. 3.11). One possibility is that CsrA and Hfq compete for binding to these sites, and their rates of association with CsrB are determined by their intracellular levels, availability and binding affinities. CsrB would then regulate not only the relative levels of free CsrA and sequestered CsrA, but also the relative pools of free Hfq and sequestered Hfq. The balancing of this sponging activity could overall fine-tune the intracellular levels of free CsrA and Hfq (Fig. 3.23 A). To elaborate on the scenario that Hfq could directly interfere with CsrB – CsrA binding, I integrated this hypothetical regulation with the known direct link between the two master regulators. CsrA is known to repress *hfq* translation (Baker et al., 2007). Therefore, by sponging CsrA, CsrB indirectly activates Hfq expression (Fig. 3.23 B). Additionally, Hfq could prevent CsrB-mediated repression of CsrA activity – thus forming a negative feed-back loop that fine-tunes Hfq expression (Fig. 3.23 B). Noteworthy, this loop would

have consequences on all CsrA targets expression. Another possibility is that CsrB is involved in Hfq-mediated RNA-RNA interactions. The latter is supported by the discovery of a large number of CsrB-sRNA interactions (Fig. 3.18). I am particularly intrigued by the CsrB and Spf sRNA interaction. These sRNAs have antagonistic roles: Spf plays a main role in catabolite repression by repressing genes involved in central and secondary metabolism and the consumption of non-preferred carbon sources. CsrB inhibits the activity of CsrA, a global regulator that represses glycogen synthesis and catabolism, gluconeogenesis and biofilm formation (Romeo et al, 1998). Therefore, CsrB indirectly activates stationary phase processes. I cannot discriminate whether CsrB or Spf acts as the sponge, but given that the CsrA binding sites are still exposed upon Spf hybridization, I postulate that CsrB binds Spf and inhibits its function. By sponging Spf, CsrB may help relieve catabolite repression. This additional layer of control may be required because, although Spf binding to Hfq decreases towards the stationary phase, it is still one of the more abundantly bound sRNAs even in stationary phase (Fig. 2.2).

Either scenario directly connects the CsrA and Hfq regulons, which is currently known to be linked only by the CsrA-mediated repression of Hfq synthesis (Baker et al., 2007), and the McaS sRNA (Jørgensen et al., 2013; Holmqvist and Vogel., 2013).

Additional regulatory circuits that involve sRNA-sRNA interactions

The Hfq CLASH independently identified known sRNA-target interactions (Fig. 3.14) and uncovered many new targets for known sRNAs (Fig. 3.15; Fig. 3.20). Throughout all sections of this Chapter I have presented isolated cases of potential sRNA mediated regulation. Many of these circuits involve sRNA-sRNA interactions, which seem to be widespread (Fig. 3.18). Sponge-sRNA interactions have been previously described (Myiakoshi et al., 2015b; Figueroa-Bossi et al., 2009; Lalaouna et al. ,2015), but most sponges originate as RNA processing products. This led to the view that RNA sponges mediate mRNA-mRNA cross-talk, the communication between the mRNA species releasing the sRNA, and the mRNA targets of the supressed sRNA. The interactions found in the Hfq CLASH data show a new layer of sponging activity,

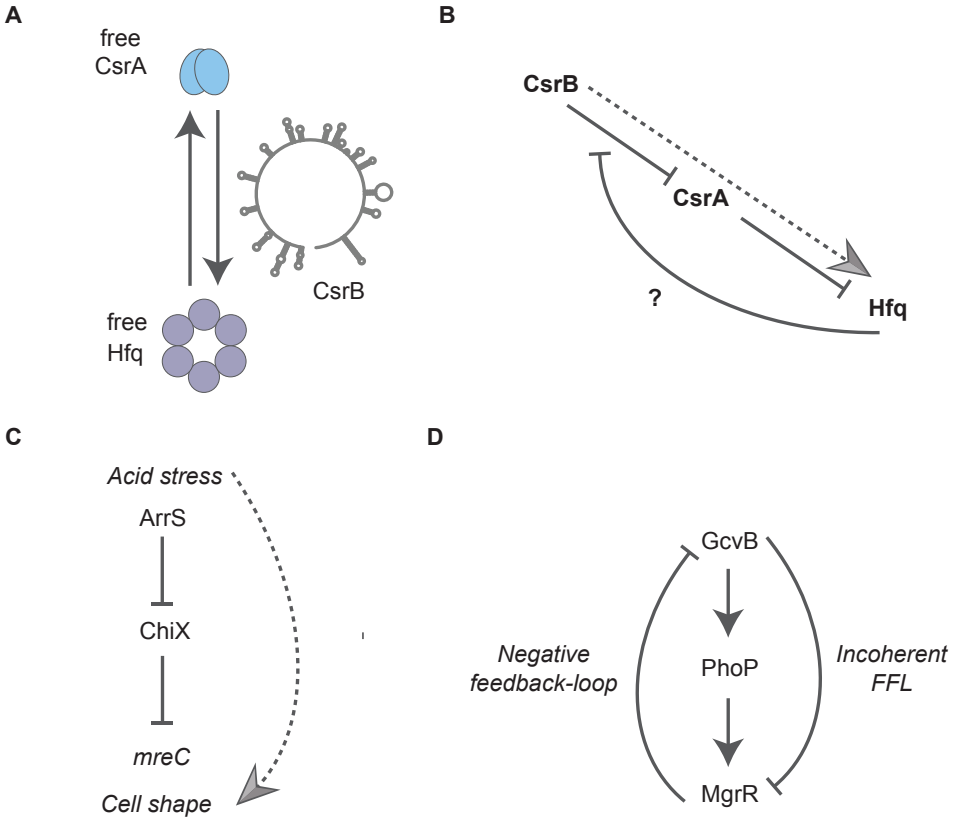

Fig. 3.23: Hypothetical models derived from observations made in the Hfq CLASH data. (A) CsrB may bind to both CsrA and Hfq *in vivo*, thus controlling the relative levels of free, active protein and sequestered, inactive protein for each of these master regulators. This ultimately would affect the relative levels between the active forms of the two proteins. (B) In the scenario that Hfq competes with CsrA for binding to CsrB, these three regulators could constitute a negative feedback-loop that fine-tunes expression of Hfq. The dashed arrow shows that expression and CsrA-sponging activity of CsrB (C) During acid stress, ArrS may sponge ChiX to derepress a set of its targets, specifically *mreC.* This regulation could link acid stress responses to regulation of cell morphology. (D) GcvB is known to activate PhoP, who in turn activates expression of MgrR. The MgrR-GcvB interaction recovered from the CLASH data suggests that either sRNA could act as a aponge depending on the cellular conditions, their relative expression level and kinetic parameters. If MgrR acts as the sponge, it could form a negative feed-back loop to keep GcvB under control. If GcvB acts as the sponge, it could repress MgrR to adsorb trnascriptional noise from PhoP, in an incoherent FFL.

whereby independently transcribed sRNAs control each other. This type of regulation has several implications: first, the expression of the sponge is not dependent on the expression of a parental RNA - it can be rapidly transcribed when needed. Second, since most of these sRNAs do not possess mature 5'-ends, the interaction need not be followed by RNase E mediated degradation and it is therefore possibly reversible. From the data I cannot determine which sRNA is the main regulator for each interaction, but one possibility is that the regulatory outcome is dependent on the relative levels

of the two sRNAs, similar to the linear-threshold model described for sRNA-mRNA pairs (Section 1.10). Moreover, since the expression of sRNAs is characterized by transcriptional noise, sRNA-sRNA interactions may help reduce it.

In one example (Fig. 3.15) I propose that ArrS might base-pair with ChiX to occlude the seed sequence for its interaction with *mreC*. This binding would not affect regulation of the *chiP* target, for which ChiX utilizes a different seed sequence (Fig. 3.15). Since ChiX is predicted to base-pair with *mreC* in the start codon region, it is likely that ChiX inhibits translation of *mreC* mRNA. Thus, ArrS might derepress *mreC*, and possibly other targets for which ChiX utilizes the same region for pairing (Fig. 3.24 C). It is possible that under different growth conditions, and depending on the stoichiometry between the two species, ChiX could sponge ArrS to repress acid-responses. Although ChiX is detectable during all growth phases (Fig. 2.2), it was found to be upregulated by low pH under specific conditions (Hayes et al., 2006). Since ArrS is involved in cell responses that tackle acid stress, the interaction between the two is likely not coincidental, and in this hypothetical model, the sponge would link acid-responses to cell shape determination (Fig. 3.23 C).

The MgrR-GcvB (Fig. 3.17) interactions may provide a crosstalk between amino acid metabolism regulatory pathways and lipopolysaccharide composition regulation. Various links between these two sRNAs have been reported, including that GcvB was shown to activate expression of PhoPQ regulon, part of which is MgrR (Moon and Gottesman, 2009; Coomaert et al., 2013). In line with this, GcvB could potentially act as a sponge to buffer transcriptional noise of MgrR, in an incoherent FFL (Fig. 3.23 D). Alternatively, MgrR could act as a sponge to feedback on GcvB sRNA activity. In this case, the RNA species that acts as the sponge is determinant of the network motif employed between these regulators (Fig. 3.23 D).

Moreover, the SroC sRNA derived from a GcvB target was shown to act as sponge for both MgrR and SroC (Acuña et al., 2016). These observations suggest that the MgrR and GcvB regulons are intimately connected at multiple levels.

The sRNA-sRNA interactions examples briefly discussed here, although still awaiting experimental validation, already indicate possible circuits and strategies the cells could use to integrate multiple adaptive responses. The discovery of so many sRNA-sRNA interactions reflects a direct crosstalk between sRNA regulons, an avenue worth exploring in more detail.

New 3'-UTR-derived sRNAs captured by CLASH

In the Hfq CLASH data, we identified ~150 mRNAs with 3'-UTRs within chimeric reads, and the consensus motif of these fragments indicates overlap with Rho-independent terminators (Fig. 3.14B). The GO-term enrichment analysis of mRNAs with 3'-UTRs within chimeras did not indicate any enrichment, which suggests that mRNAs that can give rise to sRNAs can be involved in virtually any cellular pathway.

The intersection of our 3'-UTR chimera dataset with the TIER-seq data (Chao et al., 2017), selects those 3'-UTRs that can release sRNAs by RNase E processing. Of these sRNAs, we detected by Northern Blot a sRNA derived from the 3'-UTR of *malG*, which base-pairs with RNAs of *ompC, ompA,* and *ohsC*, among other targets (Fig. 3.20 B; Section 3.11). This sRNA (MdoR; Fig. 3.22) was characterized in detail in Chapter 4.

Comparison between Hfq CLASH and RIL-seq

Throughout my analyses, I compared the sets of target interactions recovered by multiple datasets, most often between Hfq CLASH and RIL-seq (Melamed et al., 2016).
These methods are conceptually similar (Section 1.8), and the RNA-RNA interaction datasets were captured in association with Hfq for both. Each dataset covered multiple growth conditions. Both my CLASH and the RILseq studies had data on exponential and stationary phase, but the actual cell densities (OD_{600}) were different.

Ideally, a thorough comparison between the data generated by both methods should involve the application of the same bioinformatic pipelines to both datasets, however, running the RIL-seq pipeline proved to be quite challenging. As a result, I used arbitrary thresholds: each interaction had to be supported by at least two unique hybrid reads in the CLASH data and these were compared those with the significant RNA-RNA interactions reported by Melamed et al., 2016. Setting this low-stringency count threshold for the CLASH data was backed by the following rationale. First, the high-stringency conditions for purification, and the higher temperatures used in the CLASH protocol (Section 1.8; Section 2.13), ensure that only the more stable duplexes are recovered. Secondly, the first CLASH experiment I performed using the published protocol (Helwak and Tollervey, 2014) revealed few RNA-RNA intermolecular interactions. However, it contained two unique chimeras of the *malG* 3'-UTR - *ompC* 5'-UTR interaction, which I successfully validated (Chapter 4).

My bioinformatics analysis showed that there was surprisingly little overlap between the Hfq RIL-seq and CLASH data (Fig. 3.14, Fig. 3.17). This was especially surprising considering that I did not use very stringent criteria for selecting chimeras from our CLASH data. Regardless, those RNA-RNA interactions recovered from both datasets are very likely *bona fide* interactions that the lab will experimentally validate.
The poor overlap between the datasets can be explained by the differences in growth conditions. For example, the CLASH data included interactions characteristic to more growth-stage conditions (7 different OD_{600}), thus it is more likely to include more RNA-RNA duplexes formed during growth transitions from exponential to stationary phase, a phase of growth not analysed by RIL-seq. More differences in the sets of interactions could ensue from the fact that the optical densities sampled for the exponential and stationary phases of growth for RIL-seq were different from CLASH. These aspects can be explored further, by comparing the datasets for each condition separately. However, the experimental conditions employed by each method can introduce biases towards the capture of certain types of hybrids (Section 1.8). In CLASH, the stringency in interaction capture is introduced experimentally, by dual-affinity purification of Hfq-RNA complexes under completely denaturing conditions

(6M GuHCl). Our bacterial CLASH protocol also employs higher ligation temperatures (37ºC). Combined, these parameters should enrich for long duplexes, whereas short duplexes are expected to be lost. Consistent with this, the MdoR-*ompC* interaction identified here is unusually stable and can be reproduced *in vitro* in the absence of Hfq (Stuart McKellar, unpublished data).

The RIL-seq protocol however, maintains the integrity of the hexamer during purification (Melamed et al., 2016), at the cost of capturing more spurious and transient RNA-RNA interactions. These are then removed *in silico* with sophisticated statistical filters (one-sided Fisher's exact test on those pairs with at least 10 hybrid reads) (Melamed et al., 2016; Waters et al., 2017).

The question arises, at which stage it is best to introduce the stringency to obtain an accurate representation of the *in vivo* RNA-RNA interactions. I speculate the sooner in the experimental design the stringency is introduced, the more likely the output will be a more accurate representation of the *in vivo* scenario, as it is very hard to mimic the cellular conditions using statistical tests. Regardless, both approaches have been shown to work well in capturing RNA-RNA interactions, and most validations of independent observations from these data, confirmed that the methods are robust. However, both methods could be improved from a quantitative point of view - given the different strategies (and biases) for hybrid capture. Even for identical sRNA-mRNA pairs common to both datasets we expect a different number of hybrids supporting the interactions. Thus, is the relative number of hybrids for an interaction reflective of the occurrence of the interaction within cells?

Future directions

The data presented in this Chapter emphasises the importance of performing kinetic studies, as they may uncover patterns and outliers that may point to new regulatory mechanisms and rewiring of gene expression through new nodes within networks. I have identified a few exciting sRNA-mediated interactions, that show that non-canonical

modes of regulation may be more common than previously thought. However, the observations made here have to be experimentally proven to be considered valid. In addition, the biological role of the interactions should be assessed. To this end, I characterised in detail one of the findings introduced in this section, in Chapter 4. Other interactions will be validated by other students in the lab.

With respect to the analysis of hybrids, my first priority is to perform a more adequate statistical analysis of my Hfq CLASH data to remove false-positive interactions. Moreover, an informative approach for elucidating the roles of sRNA interactions in the CLASH data would be to check at which specific growth condition ($OD_{600)}$ the RNAs form a duplex.

4. A novel 3'-UTR derived small RNA controls extracytoplasmic stress responses during maltodextrin utilization in *E. coli*

4.1. Introduction

As introduced in Chapter 1, microorganisms have outstanding abilities to monitor and adapt to their environments. Under stress conditions, microorganisms rely on signal transduction systems and complex adaptive networks to sense environmental stress and to control the coordinated expression of genes involved in cellular defense mechanisms. This chapter explores further the regulation of one of the major stress responses in *E. coli,* the envelope stress response.

The *E. coli* cell envelope is a vital barrier at the forefront of interaction with its milieu: it protects cells from environmental insults, mediates information exchange and allows selective, bidirectional movement of molecules. The cell envelope is composed of an asymmetric outer membrane (OM) and an inner membrane (IM) that are separated by an aqueous peptidoglycan-containing periplasmic space (Silhavy et al., 2010). OM is rich in β-barrel proteins (OMPs such as OmpC, OmpA, OmpF; Nikaido et al., 2003) and lipoproteins (e.g. Lpp), which are some of the most stable and abundant bacterial proteins. The correct integration of these proteins into the membrane requires a sophisticated assembly machinery that includes many chaperones and proteases (e.g. Skp, DegP; Sklar et al., 2007). Given the complexity of envelope biogenesis, it is not surprising that the synthesis and assembly of OM components is tightly regulated (Ruiz et al., 2006).

Envelope biogenesis control is attained by both transcriptional and post-transcriptional mechanisms involving extracytoplasmic-stress (ESR) pathways that respond to any alterations in membrane composition (Fig.4.1; Fig.4.2).

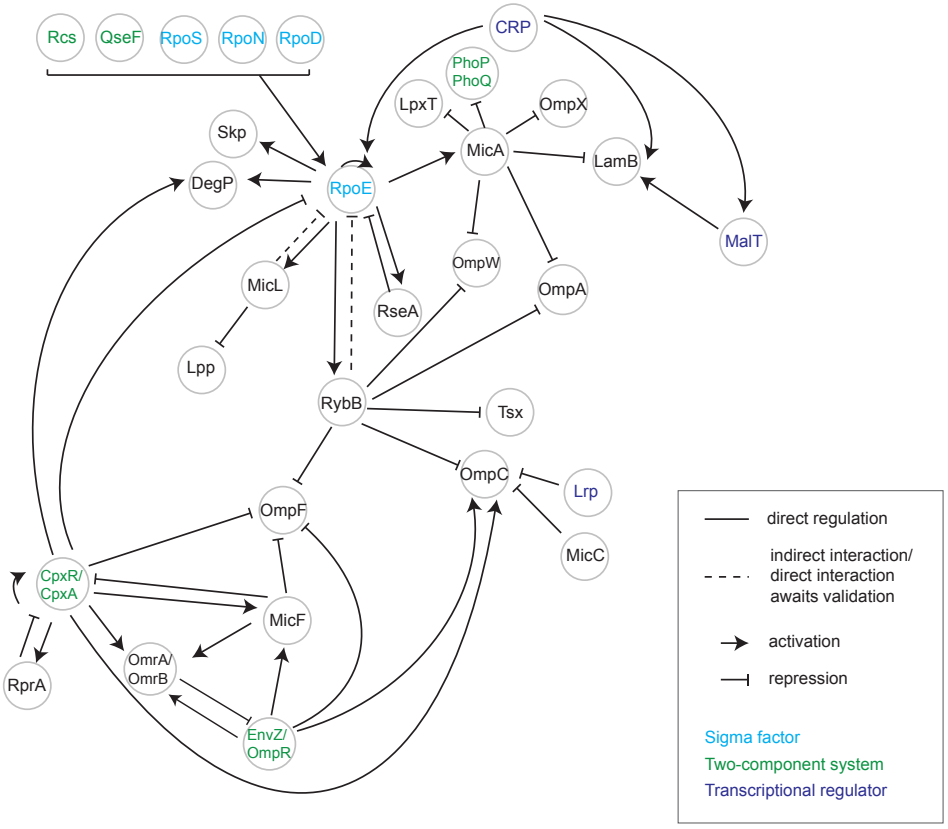

Fig. 4.1: A network of envelope stress-responsive and porin-controlling sRNAs, encompassing known and candidate interactions; the key components involved in σ^E and OMP regulation and other nodes relevant to this study are included

The alternative sigma factor σ^E (encoded by rpoE) and two-component systems like CpxR/A, Rcs, PhoPQ, EnvZ/OmpR, that respond to IM stress and elevated pH, peptidoglycan stress, lipopolysaccharide imbalances and osmotic stress, respectively (Raivio, 2014; Laubacher and Ades, 2008), control transcription of a plethora of genes with extracytoplasmic functions. As shown in Fig 4.1, most of these regulons deploy sRNAs that can either amplify the signal or dampen it. Indeed, sRNAs can act in negative feedback loops, such as the negative feedback provided by OmrA/B to OmpR or the sRNAs of the σ^E regulon (MicL, MicA, RybB) that counteract deleterious effects of σ^E expression (Fig. 4.1) (Klein and Raina, 2017; Gogol et al., 2011). Moreover, two-component system components are targeted by sRNAs from distinct pathways, as exemplified by Cpx that is targeted by CyaR and RprA (Vogt et al., 2014) and MicA-mediated repression of PhoP (Coomaert et al., 2010).

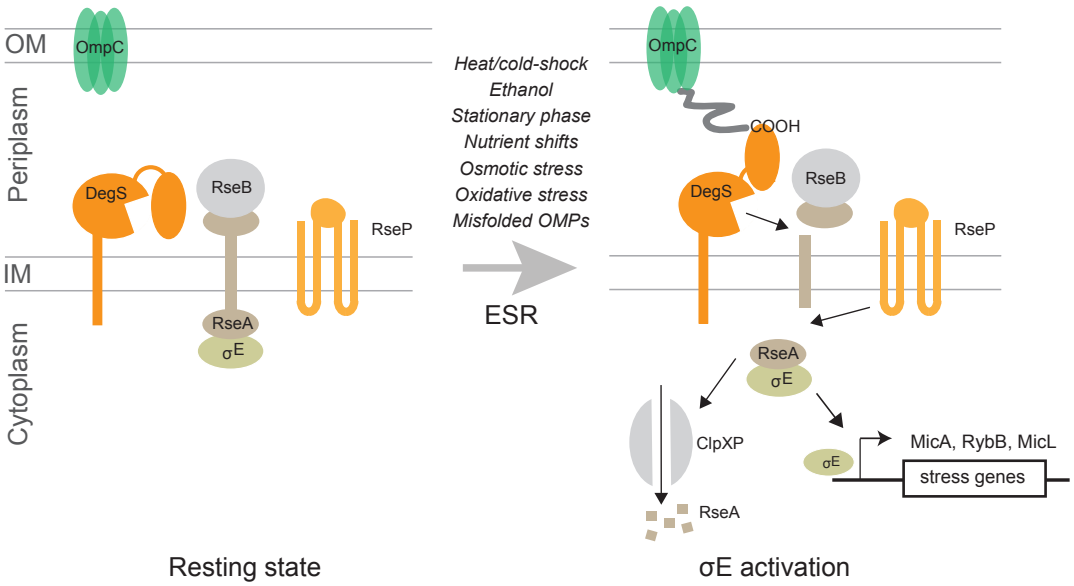

Fig. 4.2: A survey of the σ^E activating factors and proteolysis steps that release membrane-anchored σ^E into the cytosol to allow transcription of extracytoplasmic stress-responsive genes and sRNAs - MicA, RybB, MicL; OM: outer membrane, IM: inner membrane, ESR: extracytoplasmic stress response; σ^E activity is regulated by the degradation rate of its negative regulator, RseA, which holds σ^E inactive in the inner membrane; in the canonical mode of σ^E activation, degradation of RseA is initiated by DegS, and requires RseB displacement. RseP and ClpXP proteases further degrade RseA, leading to the release of active σ^E.

Evidently, the non-coding arms of envelope-related pathways not only diminish stress, but also communicate with other signal transduction systems (Klein and Raina, 2017; Grabowicz and Silhavy, 2017), allowing cells to integrate multiple responses, with profound effects on the re-wiring of gene expression.

Lack of σ^E is lethal to *E. coli* (De Las Peñas, 1997). The rpoE gene is part of the rpoErseABC operon that has multiple alternative promoters, located upstream rpoE and upstream rseA (encoding for the σ^E anti-sigma factor), respectively. Promoter choice mediates the balance between positive autoregulation and RseA-mediated negative feedback loop. σ^E-mediated responses restore membrane homeostasis upon stress imposition, and are induced by any changes in envelope composition, as signalled by misfolded OMPs, heat-shock, exposure to ethanol, pH changes, or nutrient shifts (Rowley et al., 2006; Kenyon et al., 2005). The molecular details of σ^E activation are depicted in Fig. 4.2: σ^E is activated by a proteolytic cascade that degrades the anti-sigma factor RseA and releases σ^E to the cytoplasm, where it

increases transcription of target genes. RseB binds RseA and inhibits DegS activity. RseB displacement allows DegS to cleave RseA that is subsequently cleaved by the protease RseP and a cytoplasmic portion of RseA is finally degraded by the ClpXP protease to release σ^E (Ades et al, 1999; Alba and Gross, 2004; Chaba et al, 2011). Activation of σ^E induces expression of roughly 100 genes, including genes involved in membrane repair and three sRNAs (MicA, RybB, MicL) that together downregulate synthesis of abundant porins, dozens of non-porin targets and Lpp (Mutalik et al., 2009; Rhodius et al. 2005; Papenfort et al., 2006; Udeqwu and Wagner, 2007).

OMPs have been described as one of the strongest ESR signals (Rowley et al, 2006). Even minor changes in expression and folding of these proteins can trigger σ^E activation (Fig. 4.2). Any factor that affects OMP synthesis can therefore indirectly trigger σ^E-mediated ESRs. Transcription of OMPs depends on specific cellular conditions, such as during osmoregulation or metabolic changes (e.g. OmpC repression by Lrp) (Ferrario et al., 1995; De La Cruz and Calva, 2010; Pratt et al, 1997).

The *ompC* mRNA is highly abundant and unusually stable (Emory and Belasco, 1990). However, its translation is post-transcriptionally controlled by several sRNAs, enabling fine-tuning of gene expression (Nikaido et al, 2003; Guillier et al., 2006). For instance, the strong σ^E-dependent MicA and RybB lower OMP synthesis (Fig. 4.1): MicA destabilizes both *ompA* and *lamB* (Udekwu et al., 2005, Rasmussen et al, 2005; Bossi and Figueroa-Bossi, 2009; Figueroa-Bossi et al., 2006; Gogol et. al, 2011), whereas RybB downregulates OmpC, OmpA, OmpW, and OmpF (Johansen et al, 2006; Gogol et. al, 2011; Papenfort et al., 2010). This repression prevents further accumulation of OMPs when the production of assembly factors is not sufficient to cope with folding pressures encountered during stress. This has been described as a homeostatic loop that alleviates envelope stress, and ultimately negatively regulates σ^E (Thompson et al., 2007). Further, MicC-mediated regulation inhibits OmpC synthesis by preventing translation initiation (Chen et al, 2004). Thus, a diverse repertoire of environmental stimuli assisted by a corresponding suite of sRNAs contribute to changes in porin expression, and subsequently can induce σ^E.

The envelope is also the route for nutrient intake, and its composition allows selective

passage of nutrients. Depending on the availability of the substrates, *E.coli* can switch between the expression of lower-affinity (e.g. general porins) and high-affinity transport systems for a given nutrient, as was described for glucose (Ferenci, 1996). Nutrient transport systems include components in all compartments of the envelope (Nikaido, 1994). It follows that the changes in expression of transport system proteins in response to nutrient sensing remodels the architecture of the envelope, which may trigger extracytoplasmic stress responses (ESRs).

The work I presented in Chapter 3 uncovered a possible novel Hfq-mediated post-transcriptional control of key components and regulators of envelope composition in *E.coli.* Here I describe a detailed characterization of some of these predicted interactions, which led me to uncover a novel regulatory link between nutrient availability sensing and ESRs. Among the many novel sRNA-target interactions identified by CLASH (Chapter 3), I discovered a new non-coding RNA, which I refer to here as MdoR (Fig. 3.20 B; Fig. 3.22). The data presented in this Chapter indicated that MdoR regulates the expression of general porins and σ^E regulon members in *E. coli* in response to intensification of maltodextrin assimilation. Here I propose that MdoR is processed from the *malEFG* transcript encoding IM and periplasmic components of the maltose-specific transport system (Section 4.8). Here I demonstrate that MdoR base-pairing with mRNA (and sRNA) targets directly represses OmpC expression, and indirectly promotes LamB synthesis, thereby controlling the expression of the OM component of the maltose-specific transport system. Mechanistically, LamB activation is likely a result of MdoR-sponging of the MicA repressor. I postulate that MdoR also indirectly suppresses σ^E activation triggered by changes in OM composition. I hypothesize that MdoR is therefore implicated in the reorganization of the envelope to facilitate higher efficiency maltodextrin uptake from the environment, while mitigating the deleterious effects of σ^E activation.

4.2. MdoR is a novel trans-acting sRNA derived from the 3'UTR of *malG*

One of the most interesting sRNAs discovered in the CLASH data was MdoR, whose

expression was very transient and peaked at late exponential phase (Fig. 3.22). I envisioned that the particularly transient expression of this sRNA may be correlated with a role in the adaptive responses triggered during transition from exponential to stationary phases of growth. In this chapter I describe the detailed characterisation of MdoR, with the final aim of elucidating its functional role.

MdoR overlaps the 3'-UTR of *malG*, the last gene of *malEFG* operon, which encodes three subunits of the maltose ABC transporter including the maltose-binding protein (Nikaido, 1994). A first hint that *malG* may harbour a small RNA was uncovered by Hfq immunoprecipitation experiments in *Salmonella* that showed significant Hfq enrichment at the 3'-UTR of *malG* (Chao et al, 2012). The findings were later complemented by transcriptome-wide RNase E cleavage site mapping in *Salmonella*, which indicated that in the absence of active RNase E, a fragment of the *malG* 3'-UTR fails to accumulate (Chao et al, 2017; Fig. 3.19 B). However, none of these studies followed up with experiments to identify whether the 3'-UTR of *malG* indeed contains a discrete, stable sRNA. To our knowledge, this is the first report that validates and describes this particular sRNA in detail, and the first to discover it in *E.coli.*

Corroborated, the results in Chapter 3 strongly suggest that MdoR is a *bona-fide* trans-acting sRNA that contains at least one potentially functional seed sequence. Its expression profile suggests that MdoR is mostly active at a specific growth phase transition, and the interaction with its targets is likely facilitated by Hfq. Support for the functionality of MdoR is given by the identification of many chimeric reads of MdoR with a number of genes (Fig. 3.20 B)

Prior to performing any functional studies of MdoR, I mapped the 5'-end of MdoR by primer extension (Fig 4.3 A). This revealed that MdoR is a 104 nt long RNA (Fig 4.5C). MdoR includes the entire *malG* 3'-UTR and a few nucleotides upstream of the stop codon (also see pointed arrow in Fig 4.3 B, C; Fig. 4.4 A). The predicted secondary structure of full-length MdoR suggested that the GC-rich Rho-independent terminator is contributing the most to the stability of the structure (ΔG = -16 kcal/mol). The sRNA

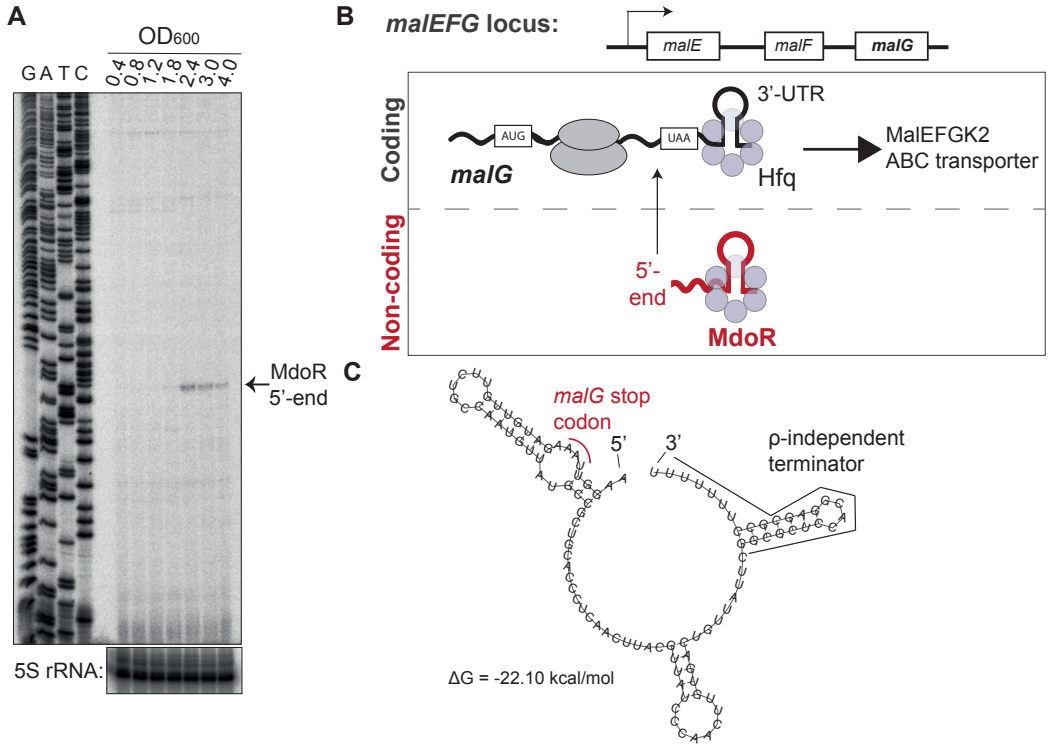

Fig. 4.3: MdoR is a 104 nt sRNA that overlaps the 3'-UTR of *malG*. (A): Primer extension analysis to map the 5'-terminus of MdoR, using *E.coli* total RNA and a RT oligo that anneals to the 3'-end region of the sRNA; G, A, T, C lanes contain the sequencing ladder; 5S rRNA was used as the loading control (B): Organization of the *malEFG* operon: the *malEFG* locus encodes a coding component that produces the MalEFGK2 ABC transporter and a non-coding component, MdoR; MdoR shares its Rho-independent terminator with that of the most downstream cistron, *malG*; Hfq binds to the Rho-independent terminator; the arrow indicates the relative localization of MdoR 5'-terminus in the vicinity of *malG* stop codon (C): *In silico* prediction of full-length MdoR secondary structure (RNAfold; Hofacker, 2003); the *malG* stop codon and the Rho-independent terminator are highlighted with red and black, respectively; the minimum free energy of folding is noted below the structure.

seed is predicted as part of weak alternative secondary structures that could become exposed by RNA rearrangements that could be chaperoned by Hfq. The CLASH data showed that Hfq binding was greatly enriched at the GC-rich stem of the Rho-independent terminator of MdoR (Fig 3.22 C). Thus, like the majority of *E.coli* trans-encoded sRNAs, MdoR is likely to be stabilized by Hfq binding at the Rho-independent terminator that it shares with *malG*, and Hfq may be required for the base-pairing and regulation of its targets.

No potential promoter sequence was identified anywhere within the upstream *malG* coding sequence, a first indication that MdoR may not emerge as an independent transcript, but could accumulate as a by-product of *malG* or *malEFG* transcript

ribonucleolytic processing, a more recently described mechanism for sRNA biogenesis (Myiakoshi et al. 2015a). The latter hypothesis was tested and is discussed in Section 4.4.

4.3. MdoR contains conserved and variable regions

Since sequence conservation among bacterial species is often indication of preservation of function, I performed a conservation analysis of MdoR and of the corresponding base-pairing regions of some of its targets (Fig 4.4 A; Fig 4.4 B). The alignments and distance tree based on percentage-identity revealed that the full-sequence of MdoR is not perfectly conserved among enteric bacteria, especially when a broader range of species are considered (Fig 4.4 C). Many *E.coli* sRNAs are very well conserved (e.g. MicC; Chen et al., 2004), however, functional sRNAs that do not display robust sequence conservation have been reported (Skippington and Ragan, 2012). It is possible that MdoR, like these sRNAs, is of a more recent evolutionary origin.

MdoR homologs of analysed species appear to be more highly conserved at the regions that are involved in the predicted base-pairing (sequences below the black lines in Fig. 4.4 A) as well as in the Rho-independent terminator region.

In *E.coli*, MdoR base-pairs with *ompC* utilising two different sites: a shorter base-pairing region separated by a short bulge structure from a downstream, more extended interacting region (sequence in the red panel in Fig. 4.4 A with the sequence in the green panel in B). However, given the higher conservation, GC content and more-extended length of the more downstream base-pairing site (second red box in Fig 4.4 A), I predict that this region primarily contributes to the interaction between MdoR and *ompC*, whereas the variable regions may help to further stabilize the interaction.

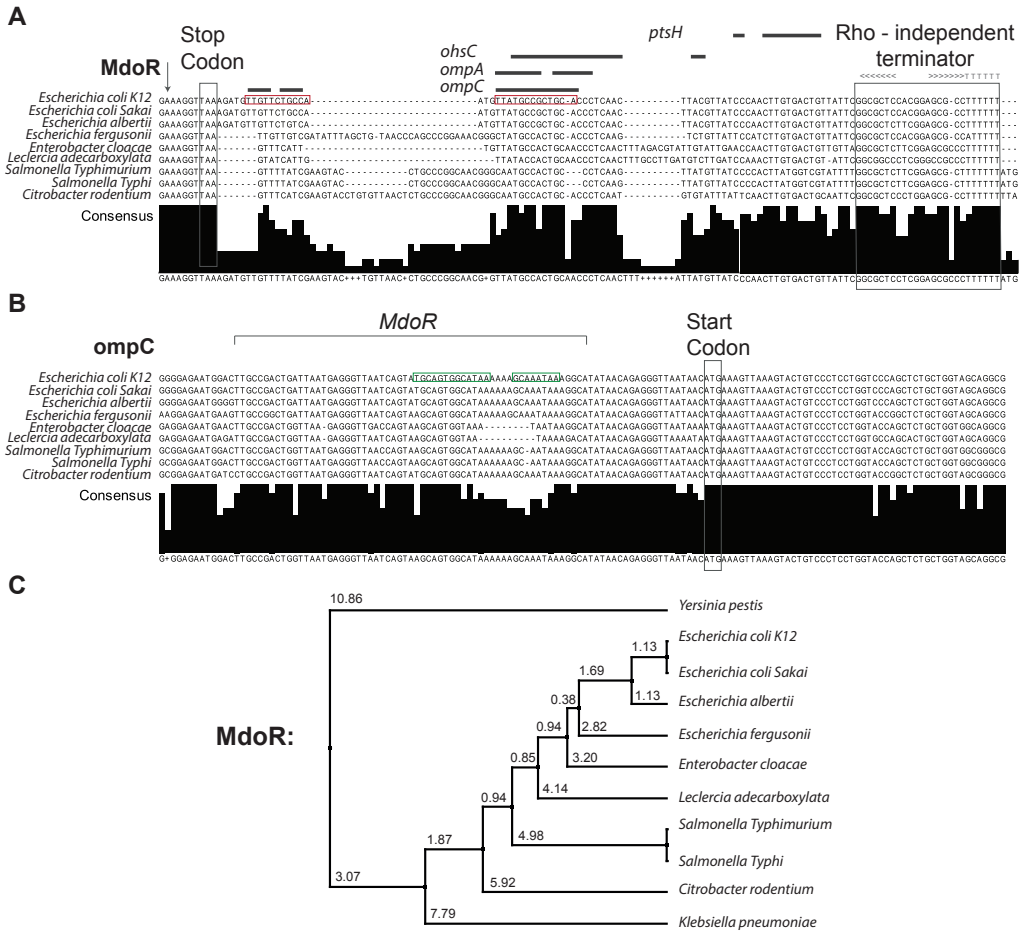

Fig. 4.4: MdoR base-pairing ability with *ompC* is preserved in enteric bacteria. (A): Sequence conservation analysis of MdoR in several Gram-negative bacteria species (Mafft algorithm with defaults); the arrow indicates the 5'-end of MdoR; the *malG* stop codon and Rho-independent terminator sites are highlighted; the horizontal black lines indicate the base-pairing regions of MdoR with *ptsH* (top), OhsC, *ompA* and *ompC* (bottom) in *E.coli* as predicted by CLASH followed by folding (RNACofold); the red boxes highlight the exact base-pairing sites with *ompC* for the species corresponding to the each sequence track (B): Sequence conservation analysis of *ompC* as in (A); the green boxes highlight the seed region for MdoR base-pairing for the corresponding species. (C): MdoR phylogenetic tree showing the distance (indicated by the values above each branch) among enterobacteria species, calculated based on percentage identity.

4.4. RNase E is responsible for both biogenesis and turnover of MdoR

Next, I asked how MdoR is generated. In bacteria, sRNAs that overlap with 3'-UTRs of mRNA transcripts generally emerge by one of two mechanisms (Section 1.7). An sRNA can either be transcribed from an internal promoter that overlaps a coding gene (parallel transcriptional output) or by processing of a mRNA precursor that is in most cases mediated by endoribonuclease E (RNase E) (Chao et al., 2012; Chao et al,

2017; Myiakoshi et al., 2015). Since I could not find an internal promoter for MdoR within the *malG* coding sequence, I hypothesised that MdoR is generated by decay of the *malG* or the entire *malEFG* operon transcript. Because MdoR is bound by Hfq at the strong Rho-independent terminator, it half-life is likely much longer than other parts of the *malG* transcript. This is highly reminiscent to what was described for other 3'-UTR-derived sRNAs, such as CpxQ (Chao and Vogel, 2016), and RaiZ (Smirnov et al, 2017) in *Salmonella*.

A close inspection of the Northern blot data of the malG 3'-UTR revealed that the full-length *malG* and *malEFG* transcripts roughly follow the same expression profile as MdoR (Fig 4.5 B): both peak at OD_{600} of 1.8 and disappear at OD_{600} of 2.4, whereas a small fraction of MdoR remains stable at OD_{600} 2.4, perhaps due to stabilisation by Hfq (Fig 4.5). Additionally, I identified shorter *malG* 3'-UTR-containing fragments of intermediate length between *malG* and MdoR that are most likely degradation intermediates. The presence of these intermediate species suggests that the *malEFG* transcript is undergoing serial ribonucleolytic cleavage steps that ultimately degrade *malG*, except its 3'-UTR, which is protected by Hfq. Hfq binding occludes RNase E cleavage sites (Moll et al., 2003), thus stabilizing the 3'-UTR containing 104 nt fragment that constitutes MdoR.

In order to test whether the identified intermediates (Fig. 4.5 A) and MdoR are degradation products, I assessed the phosphorylation state of their 5'-termini. As mentioned (Section 1.7), primary transcripts bear a 5'-triphosphate (5'-PPP) at the 5'-end, whereas processed transcripts have a characteristic 5'-monophosphate end (5'-P). To determine which type of 5' terminus MdoR has, I treated total RNA extracted from cells at OD_{600} 1.2 and 1.8, conditions in which MdoR is most abundant, with Terminator 5'-Phosphate Dependent Exonuclease (5'-P Exo). This exonuclease only degrades RNA species with mono-phosphate 5'-ends. I hypothesised that if MdoR is a degradation product, it should have a monophosphorylated 5'-end, making it susceptible to 5'-P exo degradation. Consistent with this idea 5'-P

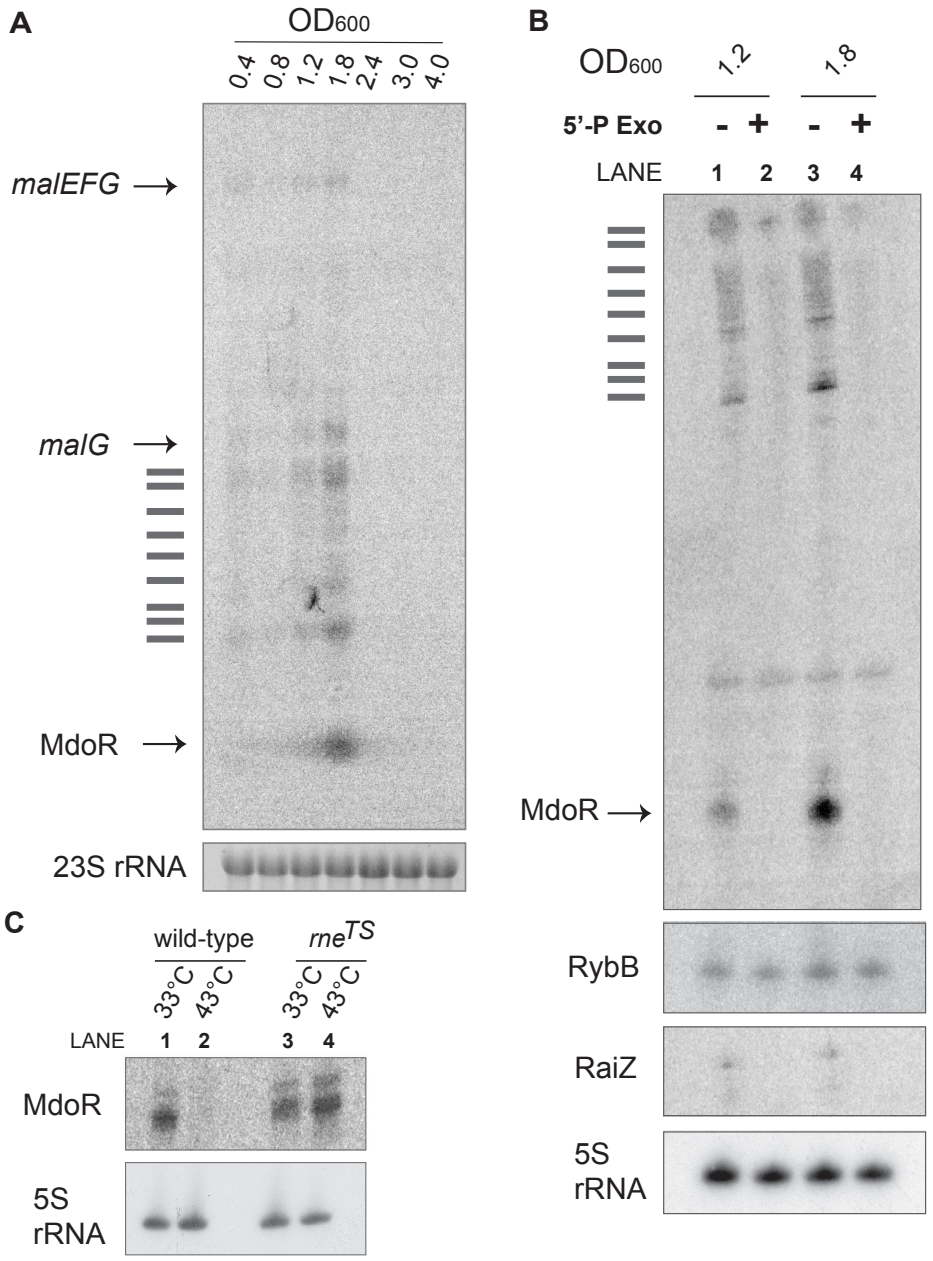

Fig. 4.5: MdoR 'life-cycle' in *E.coli*. (A): Northern blot using total RNA from *E.coli* harvested at different cell densities (OD_{600}) probed with an oligo antisense to *malG* 3'-UTR; 23 S rRNA was used as the loading control; the identity of the bands is indicated at the left of the panel; horizontal bars indicate degradation intermediates (B) Northern blot using total RNA from cells at indicated OD_{600} with (+; lanes 4 and 4) or without (-; lanes 1 and 3) 5'-Phosphate-Dependent Exonuclease treatment; 5'-P Exo only degrades RNAs with processed 5'-ends; the top panel shows the Northern probed for MdoR; The absence of a band indicates 5'-P Exo-mediated degradation; RybB was probed as a negative control; RaiZ as a positove control and 5S rRNA was probed as the loading control. (C): Temperature-dependent RNAse E *in vivo* inactivation in *E.coli.* The cultures were grown to OD_{600} 1.5 and shifted to 43 °C for 30 min (lanes 2 and 4), total RNA extracted and probed for MdoR; 'wild-type' denotes an *E.coli* MG1655 strain, whereas '*rne*^{TS}' denotes an *E.coli* N3431 strain *(rne*3071 ts). The controls were left growing at 33 °C for 30 min (lanes 1 and 3).

Exo treatment showed that MdoR was completely degraded in the presence of the enzyme (Fig. 4.5 B, lanes 2 and 4), indicating that MdoR indeed bears a 5'- mono phosphate. As a negative control, I analysed RybB, sRNA that is generated by an independent promoter (Papenfort et al, 2006; Johansen et al, 2006). As this sRNA has a 5'-PPP, it should be a poor substrate for the nuclease. Indeed, RybB levels were unaffected by the treatment with 5'-P Exo. RaiZ, a sRNA known to be generated by RNAse E processing, thus with a 5'-P, was degraded in the presence of 5'-P Exo.

Notably, not only MdoR was degraded by 5'-P Exo, but also *malG* and the degradation intermediates (Fig 4.5 A and B; horizontal bars), supporting a mechanism by which the full-length polycistronic RNA is undergoing decay that is initiated at a site in the upstream *malEFG* region. At this point, however, the ribonuclease responsible for these processing events in *E.coli* is unknown. In most described cases involving sRNA biogenesis from 3'-UTR, the ribonuclease responsible for the turnover of mRNA precursors is RNase E (Myiakoshi et al, 2015a; Chao et al, 2017).

In order to test whether MdoR is generated via RNase E degradation of the *malEFG* transcript, I used a temperature-sensitive RNAse E *E.coli* strain (*rne*^{TS}) (Apirion, 1978). This strain expresses an RNase E mutant (*rne*3071 ts allele) that causes inactivation of RNase E upon shift to high-temperatures (43-44°C) (Fig. 4.5 C). At 33°C *rne*^{TS} strain has normal RNase E activity. I rationalised that if RNase E is the ribonuclease required for MdoR biogenesis, temperature inactivation of RNase E in the *rne*^{TS} strain should prevent accumulation of MdoR. Therefore, I expected lower levels of MdoR in the *rne*^{TS} strain at 43°C (lane 4) compared to 33°C (lane 3), whereas in a wild-type strain MdoR levels should be similar at both temperatures.

Surprisingly, the levels of MdoR dropped at 43°C (Fig. 4.7 C, lane 2) even in the wild-type strain (*E. coli* MG1655). However, in the heat-shocked *rne*^{TS} strain, MdoR was stabilized. This suggests that normally heat-shock triggers the degradation of MdoR.

Given that MdoR is degraded 'by default' even in a wild-type background at 43°C,

the expectations for the changes in MdoR expression in a mutant RNase E strain still needs to be confirmed. Therefore, if RNase E mediates both MdoR generation and turnover, I expect that after the programmed inactivation of RNase E, MdoR levels should be approximately the same as before the shift (*rne*^{TS} at 33°C), with the steady state at 33 C determined by the *in vivo* rates of biogenesis and degradation.

I speculate the RNase responsible for MdoR turnover is RNase E. This is supported by the behaviour of the *rne*^{TS} strain upon heat shock: MdoR is stabilized at 43°C (Fig. 4.7C, lane 4) due to the impaired mutant RNase E ability to cleave its substrates. The temperature-shift RNase E inactivation is informative with respect to the fate of MdoR, however, it cannot provide definite evidence that RNase E is also the key player that releases MdoR from its *malEFG* precursor, especially that a slight increase in MdoR levels in the *rne*^{TS} mutant was observed at 43°C (Fig. 4.4 C).

Nonetheless, if RNase E is indeed involved in MdoR accumulation, the slight increase in MdoR levels at 43°C in the *rne*^{TS} strain may offer insights into the dynamics of MdoR generation and turnover in a wild-type phenotype background (both *rne*^{TS} at 33°C and MG1655) - namely, shortly after biogenesis, a small fraction of the MdoR pool is degraded by RNase E. This could be due to the fact not all *malG* 3'-UTRs are bound by Hfq.

Noteworthy, I have reasons to believe that our *rne*^{TS} strain (obtained from CGSC, The Coli Genetic Stock Center) may not work entirely as was observed in other studies (Guo et al, 2014; personal communication with Jai Tree). For instance, a notable accumulation of 9S rRNA, the 5S precursor, could not be identifed after 30 minutes of heat-shock. However, studies performed in a *Salmonella rne*^{TS} strain (Chao et al, 2017) suggested that the *malG* 3'-UTR release from its precursor does require RNase E. MdoR 5'-end is indeed located in a predicted single-stranded region containing two adenosines (Fig. 4.3 C; Fig. 4.4 A).

4.5. MdoR is predicted to interact with its targets using two seed regions

I have thus far established that MdoR is a trans-acting sRNA that emerges from a longer coding transcript by ribonucleolytic processing (Section 4.4). This section is the first of a number that delve into the functionality and physiological relevance of MdoR. The obvious first step into unravelling the biological significance of any sRNA is to identify its targets.

The CLASH data showed that MdoR interacts with a few mRNA targets in *E.coli* (Section 4.2). The majority of these interactions were supported by many (>= 5) hybrids, and were also found in replicate experiments. Therefore, I consider these to be real MdoR targets. To substantiate these results, I employed a bioinformatic prediction of its targetome. In recent years, computational predictions of sRNA-mRNA interactions on the genomic scale have proven to be immensely useful in the characterization of new sRNA species. Indeed, bioinformatic prediction analyses pose clear advantages. Besides the low-cost and the high-throughput of the methodology, the advantages expand to a purely biologically-relevant reasoning. Specifically, computational predictions may reveal additional targets of a sRNA that cannot be identified by experimental approaches simply due to the limitations posed by the nature of experimental work: only a limited number of growth conditions can be employed and analysed at a time. From our experience, the patterns of sRNA expression are very fine-tuned to the growth conditions: not only it is extremely challenging to find the condition at which a sRNA is maximally expressed, but it is even more difficult to identify the conditions during which a sRNA is regulating a particular subset of its targets. This is where computational predictions of sRNA-mRNA do fill a gap in the characterization of a novel sRNA: they identify favourable sRNA-mRNA interactions in a virtually infinite number of possible conditions.

For MdoR target prediction, I employed one of the most advanced tools available for transcriptome-wide sRNA-mRNA computational mapping, CopraRNA (Comparative prediction algorithm for small RNA targets, Wright et al, 2014; Wright et al., 2013).

Three homologous MdoR sequences (*E.coli, Enterobacter, Shigella)* were analyzed with CopraRNA. MdoR is predicted to have 45 targets with a combined CopraRNA p-value below 0.01, set by the developers as the highest-confidence detection threshold (Wright et al, 2013). Table 4.1 shows the top 25 CopraRNA predicted targets. MdoR is predicted to interact with targets that cover a broad spectrum of cellular functions, ranging from sugar-uptake systems, enzymatic reactions, transport to cell division regulation. The CopraRNA prediction pipeline automatically incorporates post-processing methods, such as functional enrichment analysis. However, the functional enrichment analysis using a recommended threshold of DAVID enrichment score > 1.3 (Huang et al., 2009) showed no significant clusters among MdoR predicted targets with p <0.01.

Table 4.1: Top CopraRNA hits* in *E.coli, Shigella dysenteriae* and *Enterobacter cloacae*

Rank	Gene name	Annotation	Hybridization Energy (kcal/mol)	Position mRNA	Position ncRNA
1	*caiC*	putative crotonobetaine/carnitine-CoA ligase	-53.1	168 -- 226	3 -- 71
2	*ppc*	phosphoenolpyruvate carboxylase	-44.74	157 -- 216	4 -- 65
3	*ptsI*	PEP-protein phosphotransferase of PTS system (enzyme I)	-37.75	179 -- 223	4 -- 66
4	*zraR*	DNA-binding response regulator in two-component regulatory system	-39.8	70 -- 107	6 -- 41
5	*mraZ*	RsmH methytransferase inhibitor	-54.7	155 -- 220	4 -- 66
6	*trpR*	transcriptional repressor tryptophan-binding	-35.8	225 -- 246	4 -- 33
7	*mntH*	manganese/divalent cation transporter	-33.8	192 -- 232	27 -- 65
8	*yggL*	DUF469 family protein	-41.7	160 -- 206	25 -- 65

9	*rplW*	50S ribosomal subunit protein L23	-38.9	182 -- 220	3 -- 44
10	*rseA*	anti-σ^E factor	-36.6	176 -- 216	7 -- 54
11	*ampD*	murein amidase	-36.8	83 -- 116	2 -- 37
12	*yeaH*	UPF0229 family protein	-46.1	165 -- 226	3 -- 70
13	*ligT*	2'-5' RNA ligase	-33.4	84 -- 126	6 -- 46
14	*yghD*	putative membrane-anchored secretion pathway M-type protein	-42.5	155 -- 191	12 -- 39
15	*yidB*	DUF937 family protein	-35.5	253 -- 295	24 -- 71
16	*yoaH*	UPF0181 family protein	-32.73	264 -- 293	26 -- 72
17	*norV*	anaerobic nitric oxide reductase flavorubredoxin	-31.6	219 -- 258	32 -- 71
18	*yhbE*	EamA family IM putative transporter	-39.5	190 -- 225	14 -- 54
19	*yodD*	uncharacterized protein	-51.67	177 -- 251	4 -- 72
20	*ompC*	outer membrane porin protein C	-37.23	135 -- 186	7 -- 40

*Predicted targets with CopraRNA pval <0.01 and FDR <= 0.5

CopraRNA prediction analysis cross-validates some of the targets identified by CLASH (minimum hybrid count = 2). As many as 6 overlapping targets (*ompC, ptsH, ptsI, elaB, ftsI, murE*) were identified in both datasets when CopraRNA predicted targets with p values below 0.05 were considered (Fig. 4.8 C). Both methods consistently indicate that MdoR is base-pairing with ompC in the same regions of its 5'-UTR, providing additional support that MdoR-*ompC* mRNA interaction is phylogenetically conserved.

The combined density plot of the relative frequency of MdoR nucleotide position in the predicted MdoR-mRNA interactions with p <0.01 (Fig. 4.6 A) indicates that MdoR may be utilizing two interacting domains for target recognition, a 'primary' seed and a 'secondary' seed. Both regions are predicted to base-pair with the top CopraRNA hits simultaneously, rather than alternatively, suggesting they have an additive effect

in supporting the interaction. In all cases, predicted-base pairing interactions never involved regions overlapping the Rho-independent terminator of MdoR, supporting the idea that its secondary structure must remain intact to allow Hfq binding to stabilize MdoR and facilitate duplex formation. The pattern of seed utilization may be divergent among distinct organisms, as observed for *caiC* and *ppc*: in *E.coli* the secondary interaction domain appears not to be required for recognition. Additionally, both seed

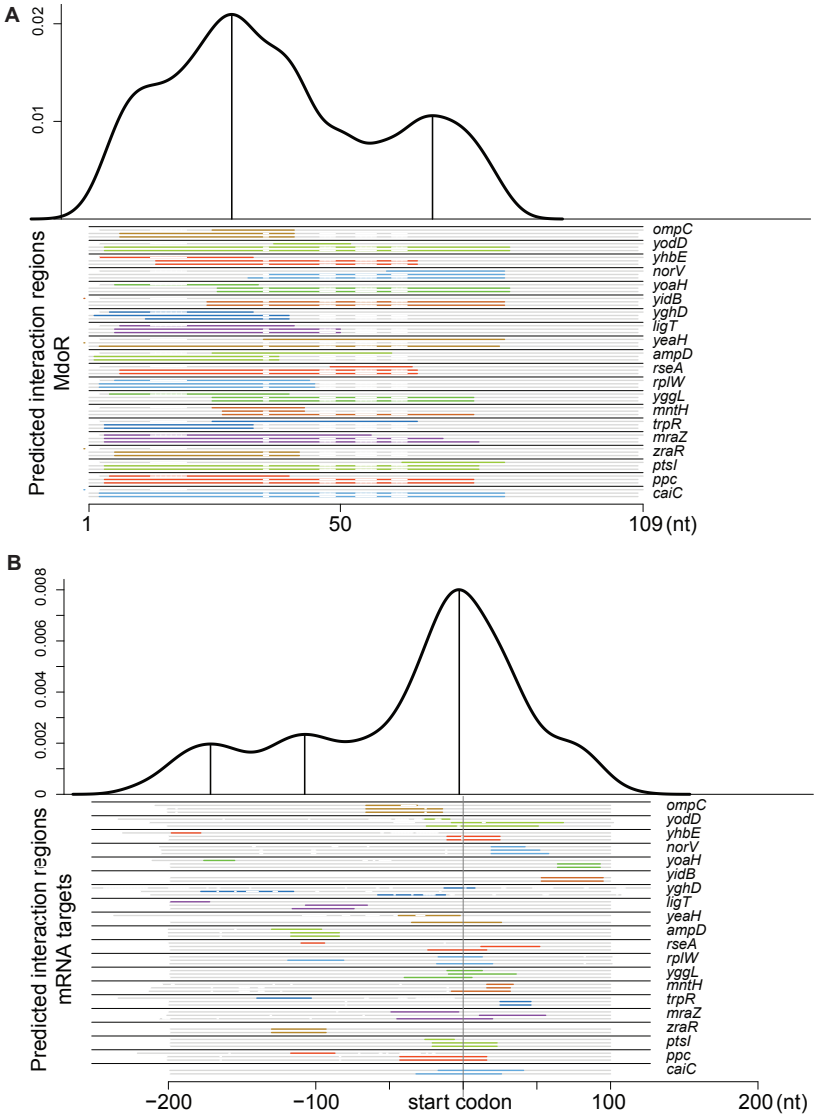

Fig. 4.6: Interaction regions within MdoR and top target mRNAs predicted by CopraRNA. (A, B) Density plots showing the relative frequency of a specific sRNA (A) or mRNA (B) nucleotide position in all predicted sRNA-mRNA interactions with a p-value < 0.01 in all considered homologs. The vertical lines indicate local maxima; the aligned regions of the homologs are shown in grey, whereas the interacting regions are shown in arbitrary colors; only the top 20 representative clusters members (Table 2) are shown in the aligned regions, with the gene names indicated on the right.

regions roughly encompass the entire length of MdoR, except its terminator, suggesting that in many cases, the entire MdoR sequence may be required for full regulatory function. It should not be excluded that the two seeds may be utilized by MdoR independently for interacting with other targets. A clear example is *ptsH*, for which the hybrids captured by CLASH point to the secondary seed as the sole interacting region (Fig. 4. 4 A).

The density plot of the relative frequency of target nucleotide position in the top CopraRNA-predicted MdoR-mRNA interactions (Fig. 4.8B) suggests that MdoR is likely to act as a de bona-fide trans-acting sRNA. Most predicted mRNA targets (e.g *ptsI, yggL, norV*) are bound by MdoR in the +/- 50 bp window relative to the start codon in all considered homologs. This is consistent with a canonical mode of sRNA-mediated regulation involving control of ribosome accessibility (Vogel and Luisi, 2011; Bouvier et al., 2008). The second most frequent interacting region within the targets is located more upstream, deep in the 5'-UTR (e.g. *zraR, ampD, ligT*), where the most commonly described sRNA-mediated regulation involves the control of mRNA stability/transcription termination. Within target mRNAs, the MdoR-interacting region appears to alternate between upstream 5'-UTR localization in one species relative to the others, indicating that different organisms may employ different modes of regulatory control. Indeed, it has been proposed that homologous sRNAs from different species may employ the same regulation and sequence, but display differential preferences with respect to target mRNA sequences (Davis et al., 2005).

Although the CopraRNA prediction analysis is very informative with respect to the assessment of regulatory potential of MdoR and may offer quick insights into its functional or even physiological role, the findings should be experimentally validated. Several predicted MdoR targets were already confirmed by CLASH. However, neither methodology is exhaustive, so additional *in vivo* screens for MdoR targets were performed (Section 4.6), followed by validations (Section 4.7).

4.6. MdoR regulates expression of outer membrane porins and key members of the σ^E regulon

In order to corroborate the CLASH data and the CopraRNA predictions, I pulse-overexpressed MdoR using a plasmid containing an arabinose-inducible promoter. I then extracted total RNA, removed riboamal RNA (see methods chapter), generated RNAseq libraries and performed a differential expression analysis (DESeq2, Love et al., 2014) on three biological replicates. The assay was set-up in a manner that would prompt minimal alterations to the system. To achieve this, the chromosomal copy of MdoR was not disrupted. I also aimed to attain maximal MdoR over-expression levels with minimal duration of induction, because a short induction time is crucial to avoid indirect regulatory effects. Additionally, an OD_{600} of 0.4 was set as the induction starting point for two main reasons. First, MdoR is endogenously expressed at this OD_{600}, thus increasing the chances of capturing the *in vivo* target repertoire relevant to a condition during which MdoR is normally active. Secondly, the level of MdoR at this OD_{600} is relatively low, such that the degree of competition of endogenous MdoR with the over-expressed MdoR variant is also kept at a minimum. However, one pitfall of plasmid-borne overexpression of MdoR is that the over-expressed variant (primary transcripts) will bear 5' triphosphates, instead of monophosphates (see Section 4.4). This may have important regulatory consequences, as outlined below.

Initially I performed a time-course experiment to optimise MdoR induction conditions. Ultimately MdoR expression was induced for a total of 15 minutes, a sufficiently-short amount of time to avoid major indirect effects (Fig. 4.7 A). As a control, I induced overexpression for the same amount of time from an empty plasmid (referred to as pBAD control). The differential expression analysis (DeSeq2, Fig. 4.7 B) identified 20 transcripts that were significantly enriched in the pBAD control data compared to the MdoR overexpression data. These transcripts are therefore very likely downregulated by MdoR (Fig. 4.7 B, Table 4.2), whereas only two are upregulated. The majority of mRNAs downregulated by MdoR are encoding for membrane-associated proteins.

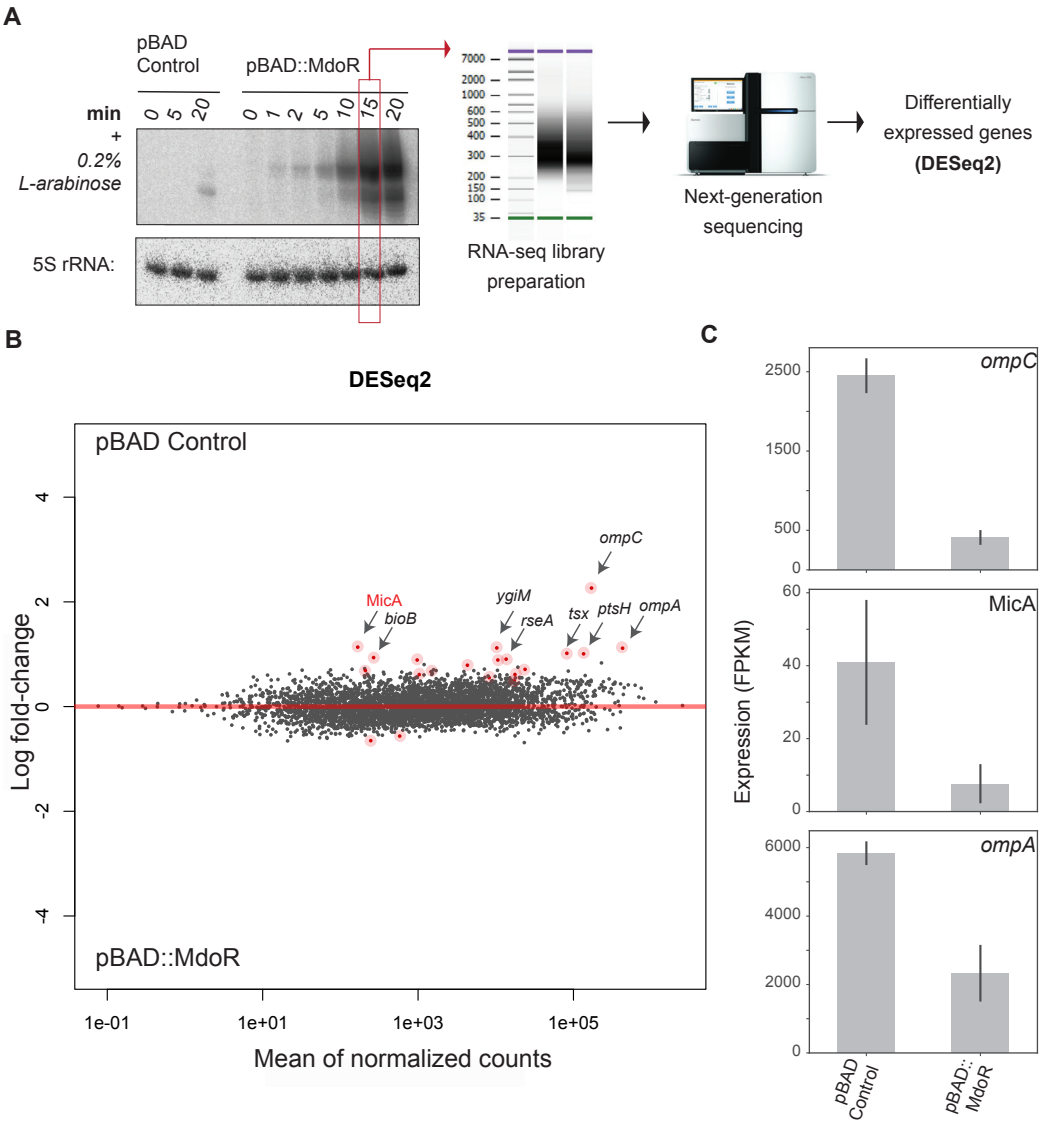

Fig. 4.7: MdoR pulse over-expression studies in in wild-type *E.coli.* (A) Time-course for MdoR overexpression conditions optimization; MdoR was induced with L-arabinose from pBAD::MdoR, whereas the pBAD Control served as an empty-plasmid control not expressing a sRNA; the 15-minute induction time was optimal and used for pulse-overexpression followed by RNA-seq library prep (bioanalyzer gel of the pooled libraries is shown) and differential expression analyses. (B) Differential expression analysis of the pBAD Control and pBAD::MdoR induced overexpression data (DESeq2 with data from three replicates): the red points with a pink halo indicate the transcripts differentially expressed in the data; transcripts with a $\log_2$ fold change > 0 were enriched in the Control data; those with a $\log_2$ fold change < 0 were enriched in the MdoR overexpression data; arrows indicate the most highly expressed transcripts in the pBAD Control data relative to the MdoR overexpression data - red font indicates sRNAs. (C) The three most-enriched transcripts in the pBAD Control data relative to the MdoR overexpression data; the y-axis indicates normalized expression (FPKM); error bars indicate standard deviations from three biological replicates.

Table 4.2: Significantly* enriched transcripts in the pBAD control data relative to MdoR overexpression data

Rank	Gene	log2(Fold change)	Gene product	Comments
1	*ompC*	2.26	outer membrane porin C	regulated by σ^E sRNAs
2	*micA*	1.14	σ^E dependent sRNA	regulates OmpA and LamB
3	*ygiM*	1.12	inner membrane protein	part of σ^E regulon
4	*ompA*	1.12	outer membrane porin A	MicA target
5	*tsx*	1.02	nucleoside-specific channel	regulated by σ^E sRNAs; OM localization
6	*ptsH*	1.01	phosphorelay protein of the PTS	-
7	*bioB*	0.94	biotin synthase	-
8	*rseA*	0.91	anti-sigma factor for σ^E	part of σ^E regulon
9	*chiA*	0.89	endochitinase	-
10	*nagE*	0.89	N-acetylglucosamine permease	N-terminus in the membrane
11	*ryeA*	0.79	sRNA	role in biofilm formation
12	*rpoE*	0.71	σ^E	part of σ^E regulon
13	*yddA*	0.69	predicted ABC transporter	IM localization
14	*chaA*	0.66	Na+/K+:H+ antiporter	IM localization
15	*wzzB*	0.61	regulator of length of O-antigen	part of σ^E regulon
16	*sixA*	0.61	phosphohistidine phosphatase	part of σ^E regulon
17	*msrB*	0.55	methionine sulfoxide reductase B	oxidative stress protection
18	*ftsZ*	0.46	cell-division protein	part of σ^E regulon
19	*yjhD*	-0.56	surface adhesin E-like protein	-

* DeSeq 2 adjusted p value < 0.05

In agreement, functional enrichment analyses (STRING, Szklarczyk et al., 2015) revealed that there was one significant pathway enrichment observed in the set of downregulated mRNAs, namely Cellular component GO:0046930, pore-complex (FDR <0.01). Additionally, there is a noticeable overlap between MdoR-downregulated genes and members of the σ^E regulon, including the *rpoE* gene itself (Rhodius et al, 2005; Table 4.3). The transcripts significantly downregulated by MdoR over-expression also included some without envelope related functions, most notably enzymes. By far the most highly downregulated mRNA during MdoR over-expression was *ompC* (~ five-fold change), supporting the idea it is a direct MdoR target. Corroborated with the CLASH data and CopraRNA predictions, this ultimately sets *ompC* apart as the MdoR target detected by all global identification approaches employed in this study, using the most stringent criteria. The accumulated evidence strongly suggests that MdoR base-pairs with *ompC* 5'-UTR *in vivo*, leading to its downregulation.

Considering the overlap between MdoR candidate targets and σ^E-regulated genes (Table 4.2), I questioned whether MdoR overexpression might have led to indirect repression of the σ^E regulon. Since the *rpoE* levels were not massively reduced (1.6-fold-change) upon MdoR over-expression, it seems more likely that the very strong repression of OmpC and other porins (OmpA, Tsx) synthesis led to an indirect relief of the σ^E response, as was proposed for other OMP-regulating sRNAs (e.g. RybB and MicA; Gogol et al., 2011; Thompson et al., 2007). MdoR over-expression also reduced the transcripts encoding for the σ^E anti-sigma factor, RseA. Since both are part of the same transcription unit, together with rseB and rseC, it is possible MdoR base-pairs to only one region of rpoErseABC polycistronic mRNA and destabilizes it. Although rseA was also a CopraRNA predicted target, the CLASH data did not indicate a region with strong Hfq cross-linking within this transcript, suggesting no trans-acting sRNA binding.Taken together, these observations indicate that MdoR indirectly represses expression of *rpoE*. Intriguingly, two of the targets that were significantly reduced upon MdoR over-expression were sRNAs (Table 4.2). The MicA sRNA (Fig. 4.7 B and C), was the second most repressed gene (~ 2.2-fold-change in the DESeq 2 data). This is a second line of evidence that MdoR is potentially capable of regulating other

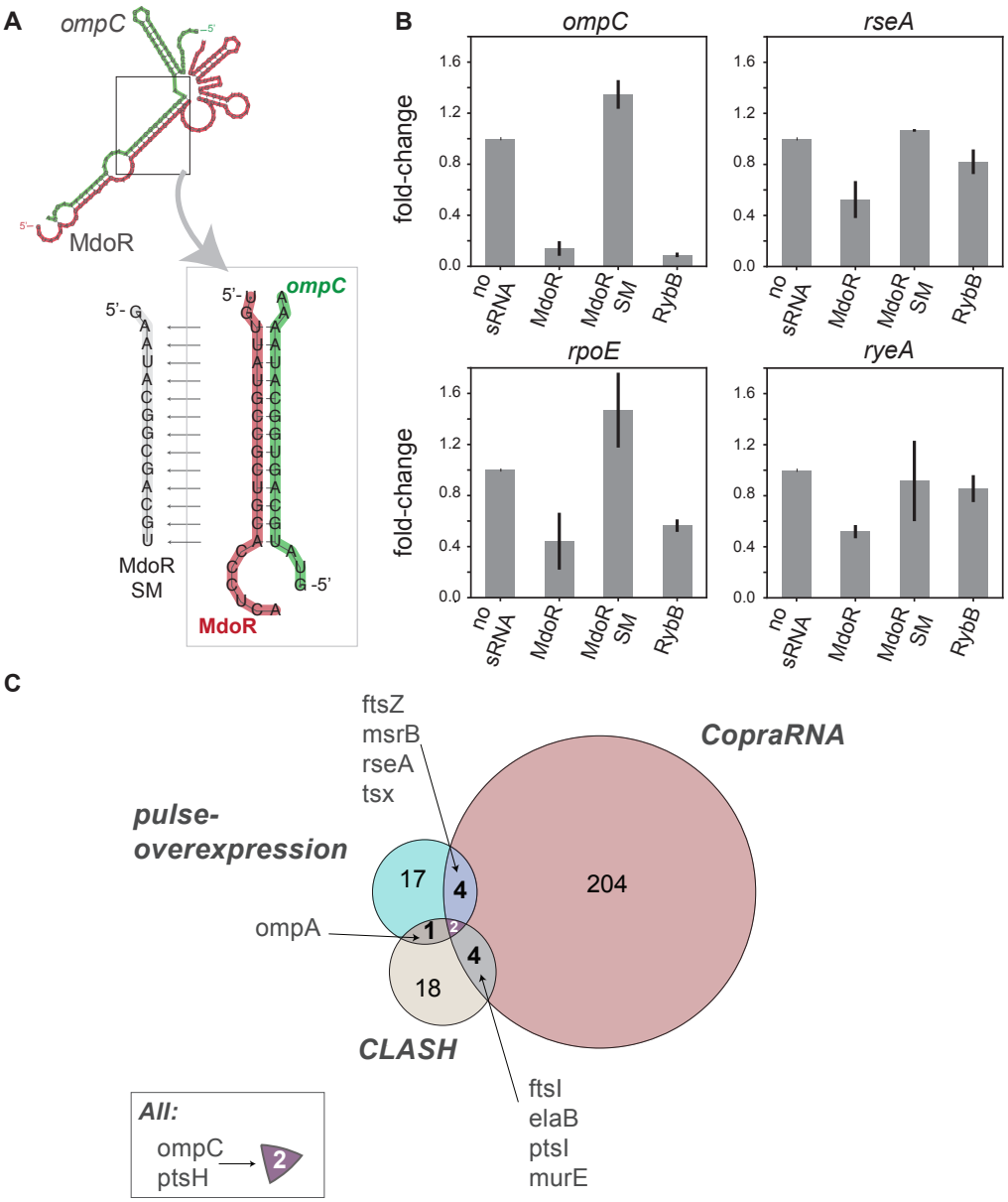

Fig. 4.8: Validation of differential expression of targets upon MdoR overexpression (A) Construct design for MdoR SM: the panel indicates and zooms in the base-pairing region within the MdoR-*ompC* duplex that was mutated; in the seed mutant (SM), the seed region of MdoR was mutated to its complement. (B) RT-qPCR analysis of several DEseq2 analysis enriched transcripts (gene names shown at the top of the plots), in the presence of sRNAs expressed for 15 minutes from plasmid-borne arabinose-inducible promoters; sRNA names are indicated below each bar; the 'no sRNA' sample contains the empty pBAD and does not overexpress an sRNA and was set as reference for fold-change calculations; the *recA* gene was used as the internal control; experiments were performed in technical triplicates; the standard error of the mean (SEM) of two biological replicates are reported as error bars; (C) Venn-diagram showing the comparision between three transcriptome-wide MdoR target identification datasets; the pulse-overexpression dataset includes all significantly down-regulated transcripts in the MdoR overexpression condition (adjusted p value < 0.05); the CopraRNA prediction dataset includes all transcripts with a combined CopraRNA p value < 0.05; The CLASH dataset includes all gene names of transcripts with at least 2 unique hybrid reads.

sRNAs, in addition to the CLASH data that indicated OhsC, another sRNA, as a target (Chapter 3). Direct base-pairing between MdoR and MicA still awaits validation, yet we are inclined to hypothesise that the downregulation is direct (see discussion below).

To validate some of the RNA-seq data, I performed quantitative RT-PCR. These experiments confirmed that MdoR over-expression reduces the levels of *ompC, rseA, rpoE* mRNAs and *ryeA* RNA) (Fig. 4.8 B).

The experiments included a control with a MdoR seed mutant (MdoR SM; see Fig.4.8 A for the design), in which the region predicted by CLASH and *in silico* folding to contribute the most to the interactions was mutated to its complement.
Pulse-overexpression of MdoR SM was no longer able to down-regulate these transcripts (Fig. 4.8 B).

It is important to note that MdoR and the σ^E –controlled RybB have overlapping target sequences in *ompC*. Therefore, I also tested whether RybB over-expression had any effect on the expression of MdoR candidate targets. As expected, a 15 min pulse-overexpression of RybB reduced both *ompC* and *rpoE* levels, but not *rseA* and RyeA, providing evidence that although these two sRNAs may have some targets in common, they are definitely not redundant.

As previously mentioned, the overexpressed MdoR variant lacks a 5'-monophosphate. The regulatory potential of an sRNA bearing a 5'-monophosphate (5'-P) is much more pronounced than its triphosphate equivalent due to preferential 5'-P-assisted cleavage by RNase E (Mackie, 1998). Thus, it is likely that MdoR works by a mechanism in which MdoR base-pairing to its target recruits RNase E and is followed by ribonucleolytic cleavage of target (and sRNA). If endogenous 5'-P MdoR indeed employs this mode of target regulation *in vivo*, I expect to have missed a number of targets or under-estimated the regulatory effect of MdoR in our pulse-expression analyses.

To conclude, MdoR pulse-overexpression caused significant downregulation of only

a handful of targets, indicating a very specific, targeted response; if any indirect effects were caused in spite of the short overexpression time, they are probably the outcome of very fast-acting indirect pathways. The immediacy of MdoR-triggered downregulation of some RNAs in *E. coli* (e.g. MicA) under such conditions may be at least as relevant in physiological terms as the proven direct base-pairing interactions (see Section 4.10).

Overall, I have employed three distinct methods for the identification of MdoR regulated-genes in *E. coli* (summarized in Fig. 4.8 C). All three point to *ompC* and *ptsH* as direct MdoR targets, whereas *ompA* is the only target common to both CLASH and pulse-overexpression analyses (Fig. 4.8 C). These methodologies are based on very different criteria, and although in some cases they cross-validate each other, they should/could be thought of as complementary methodologies.

Towards the scope of this work I will focus the biological interpretation on a select portion of the *in vivo* data, which, thus far indicate that MdoR downregulates expression of a select number of porins, indirectly represses σ^E and may direct sRNA turnover/ activity.

4.7. MdoR directly represses OmpC and OmpA expression

To demonstrate direct target regulation *in vivo*, I employed a well-established reporter system where a sRNA is co-expressed with a fusion of the target 5'- mRNA region to superfolder green fluorescent protein (sfGFP) (Urban and Vogel, 2007; Urban and Vogel, 2009; Corcoran et al., 2012) (Fig. 4.9 A). Fusions for OmpC, OmpA, Tsx, σ^E and mutant variants for OmpC and OmpA were constructed, but only OmpC, OmpA and OmpC seed mutant (OmpC SM, Fig. 4.9 B) produced stable fusions that could be analysed. A decrease in florescence relative to the control indicates downregulation, whereas an increase in fluorescence indicates upregulation. Fluorescence measurements of the translational fusions confirmed that *ompC* and *ompA* are

controlled by MdoR, whereas the GFP expression alone remained unaffected (Fig. 4.9 C.

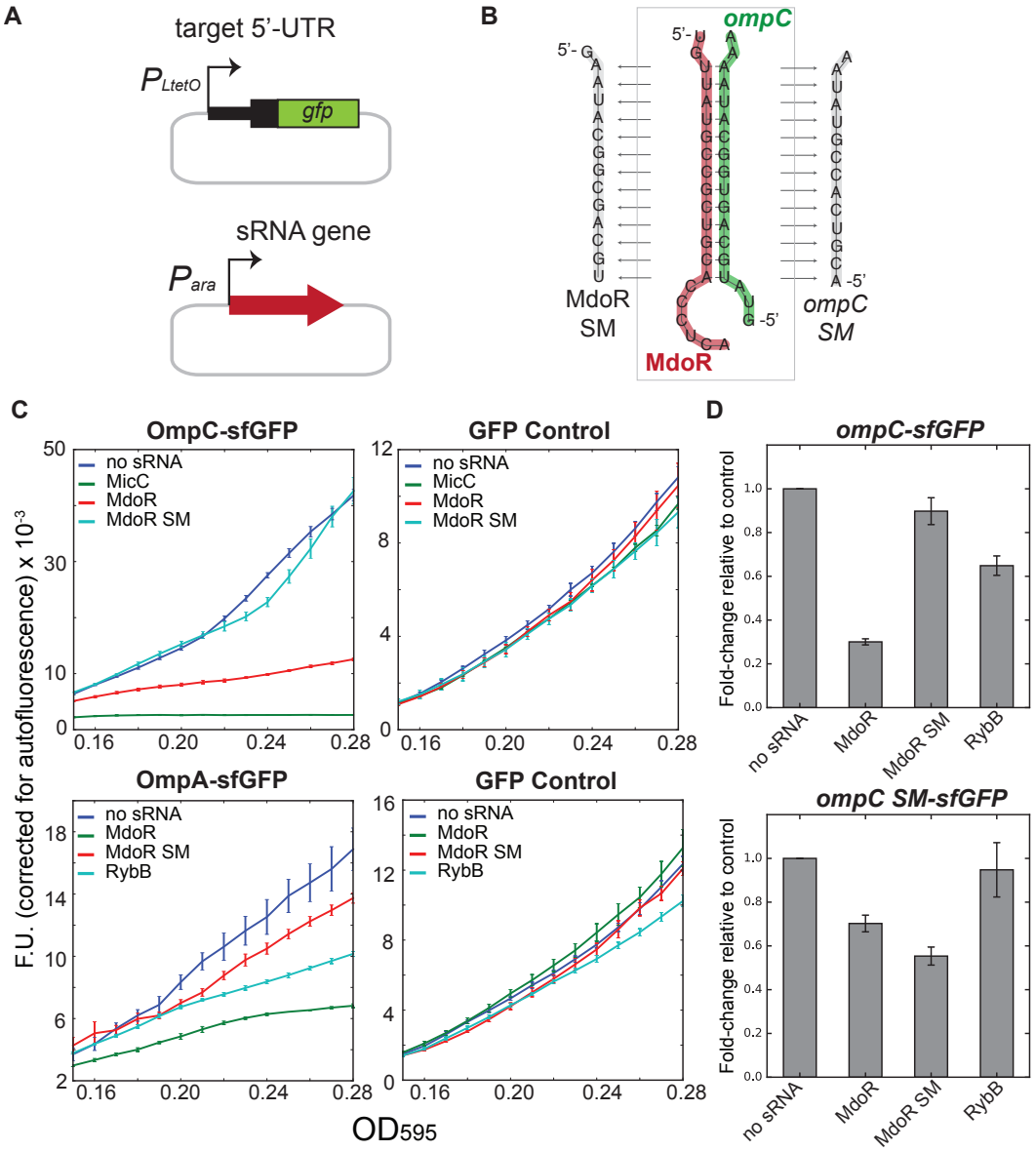

Fig. 4.9: GFP-reporter validation of OmpC and OmpA targeting by MdoR. (A) Plasmids used for the reporter assay; In *E.coli* TOP10 cells, low-copy plasmids constitutively overexpress target 5'-UTRs fused to sfGFP and high-copy plasmids overexpress the full-length sRNAs upon induction with L-arabinose. (B) Construct design for OmpC: the panel indicates the base-pairing region within the MdoR-*ompC* duplex that was mutated; in seed mutants (SM), each seed region sequence was mutated to its complement. (C) Plate reader *in vivo* fluorescence measurements of sfGFP fusions for OmpC (top) and OmpA (bottom), each with a full-length GFP Control (right), in the presence of sRNAs; the 'no sRNA' expressing strains contain the empty pBAD control plasmid; the y-axis indicates fluorescence units (F.U.); experiments were performed in technical triplicates; the means and SEM of three biological replicates are reported. (D) qRT-PCR measurements of *ompC* and *ompC SM-sfGFP* reporters; overexpression of sRNAs was induced for 20 minutes; the 'no sRNA' control was used as the reference for fold-change calculation; experiments were performed in technical triplicates; the error bars indicate SEM of two biological replicates.

As positive controls, I included RybB and MicC overexpression studies. MicC downregulates *ompC* expression at the translational level (MicC, Chen et al., 2004), whereas RybB destabilizes *ompC* and *ompA* mRNAs (Papenfort et al., 2010; Gogol et al., 2011). The results confirmed the native regulation exerted by either sRNA. In addition, The MdoR seed mutant was unable to regulate either OmpC or OmpA fusions.

Since fluorescence measurements of translational fusions cannot discriminate whether OmpC downregulation is caused by direct RNase E-mediated transcript destabilization or by translational inhibition, I also measured the reporter by quantitative RT-PCR (using sfGFP-specific primers) after only 20-minute sRNA induction (Fig. 4.9 D). As expected, reduced levels of *ompC*-sfGFP levels were observed in the presence of MdoR, but not that of MdoR SM. Next, I introduced compensatory mutations into the *ompC* 5'-UTR (*ompC* SM) to restore base-pairing with the MdoR SM mutant (Fig. 4.9 A). This partially restored MdoR-dependent regulation (Fig 4.9 D). MdoR wt partially suppressed the OmpC-GFP SM mutant.

Taken together, these results and the CLASH data strongly suggest that MdoR directly base-pairs with *ompC* 5'-UTR resulting in post-transcriptional repression of OmpC synthesis. These results place MdoR in the relatively large family of sRNAs that control porin expression in *E.coli* (Section 4.1), but MdoR has a number of distinguishing targets and features.

4.8. MdoR expression is dictated by maltose utilization needs in *E.coli*

The following sections will delve deeper into the physiological relevance of MdoR. Besides target identification and validation, I aimed to gain a better understanding of the factors that control MdoR expression. I established that MdoR is processed from the *malEFG* mRNA precursor (Fig. 4.4), and proposed a model by which MdoR is generated constitutively by RNase E processing from a fraction of the *malEFG* pool

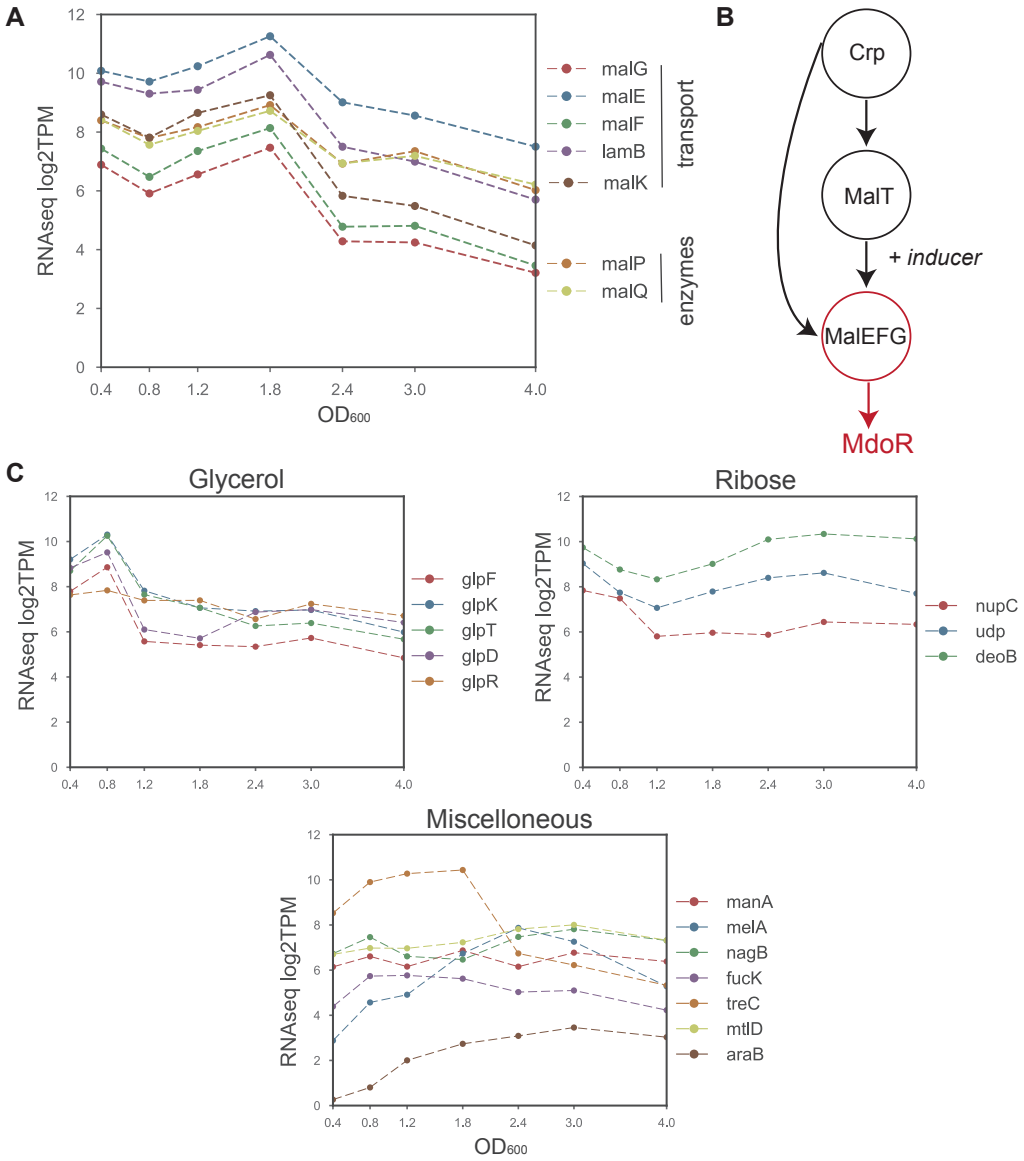

Fig. 4.10: Expression of various C-source utilization systems of *E.coli* during growth in LB (A) RNA steady-states (RNA-seq experiment) of of mal regulon genes; as normalized $\log_2$TPM (transcripts per million, y-axis); the x-axis indicates the growth-stage as optical density units (OD_{600}). (B) Feed-forward loop that leads to MdoR accumulation; MalT needs to be bound to the inducer to activate transcription of downstream genes. (C) RNA steady-states as in (A) for genes required to assimilate various substrates, indicated above panels.

of precursors. While Section 4.4 focused on gaining molecular mechanistic insights of MdoR biogenesis, this section will deal with the biological context of MdoR expression. First, MdoR is derived from a housekeeping sigma factor (σ^{70}) dependent operon. *MalEFG* has no promoter sequence for σ^E, hence, unlike RybB and MicA, two major porin-downregulating sRNAs, MdoR is not induced directly by σ^E. Since the expression profile of MdoR is dependent on and roughly recapitulates that of *malEFG*,

I rationalised that conditions that induce *malEFG* expression would by default lead to a proportional increase in MdoR expression. Therefore, elucidating the pattern of *malEFG* expression during *E.coli* growth in LB would directly provide an explanation for the emergence of MdoR. In *E.coli*, *malEFG* encodes for the translocating component of the MalEFGK2 ABC-transporter (*malF* and *malG*) and for the maltose-binding protein *(malE*), which enable the high-affinity transport of maltose and maltodextrins (α-1,4-linked oligosaccharides up to seven glucose units) from the periplasmic space to the cytoplasm (Boos and Shuman, 1998; Dippel and Boos, 2005). The maltose regulon also includes the LamB maltoporin in the outer membrane and maltose/maltodextrin catabolizing enzymes, allowing *E. coli* to feed on these substrates.

Under the growth conditions employed in this study, *malEFG* expression slowly increased from OD_{600} 0.4, until it peaked at around OD_{600} of 1.8, after which it dropped to almost undetectable levels (Fig. 4.5 A). During my search for the factors causing this particular pattern of expression, I questioned whether *E.coli* might be utilizing maltose/maltodextrins most intensely at the transition between exponential and stationary phases of growth (OD_{600} 1.8). However, the mal regulon can be induced by other physiological conditions, such as during osmotic stress (Boos and Schuman, , 1998). In those conditions the post-transcriptional regulation of *malEFG* may be very different, and MdoR may not even be generated.

In order to confirm that MdoR accumulates during maltodextrin utilization and to understand why its expression peaks at OD_{600} 1.8, I researched more closely the broader context of carbon-source (C-source) utilization by *E.coli* in LB. LB is a complex media, with very low glucose content (less than 100 uM, Sezonov et al., 2007), and other mixtures of sugars up to 0.16 % total carbohydrate content (Baev et al., 2005). Chemical analysis of complex media is not reliable with respect to the availability of nutrients, especially that many growth substrates in these media are oligomeric, and the extent of polymerization determines availability. As the tools to directly measure maltodextrin influx are not available, I used the growth-curve gene expression data (RNA-seq; Chapter 3) as readout of substrate utilization in *E.coli* (Fig. 4.10 A and

C), an approach also employed in previous studies (Baev et al, 2005). Glucose, the preferred carbon source is rapidly depleted from the medium by the time cultures reach OD_{600} of ~0.3 (Sezonov et al, 2007). At densities lower than 0.3, catabolite repression systems are active, so *malEFG* expression is expected to be low (Boos and Shuman, 1998).

When cell densities reach levels higher than OD_{600} of 0.4, Crp-cAMP is active and *E. coli* cells are already sensing and scavenging other carbon sources. The RNA steady-states suggest that at OD_{600} 0.4, *E. coli* is assimilating a few alternative carbon sources in parallel, among them glycerol, maltose/maltodextrins, D-mannose, ribose etc. (Fig. 4.10 C). Only for a few of them I noticed a sequential induction indicative of preferential utilization at a given time (e.g glycerol at OD_{600} 0.8 and maltodextrins at OD_{600} 1.8).

As shown in Fig. 4.10 A, the RNA-seq data indicates that both maltodextrin transport systems and maltodextrin-metabolizing enzymes have almost identical expression profiles, as they all peak at OD_{600} 1.8, consistent with a highest-intensity maltodextrin assimilation stage in *E. coli.* The sustained increase of *lamB* and *malEFG* expression correlates with an increased need for transporters as maltodextrins are being depleted from the medium.

Maltodextrin metabolism intensifies with elevated maltodextrin influx, as indicated by increased expression of mal regulon enzyme-encoding genes, such as *malQ* and *malP* (Fig. 4.10 A). This rewiring of gene expression is achieved by both transcriptional and post-transcriptional regulation. The genes of the mal regulon are under direct transcriptional control by Crp and MalT; MalT requires both the presence of maltotriose inducer (Richet et al., 1989; Raibaud et al.,1987) and Crp for activity, thus the mal regulon is controlled by a coherent FFL (Section 1.9) (Fig. 4.10 B).

The drop in *malEFG* steady-states after OD_{600} 1.8 can be explained by the complete depletion of maltodextrins from the environment at this cell density: in the absence

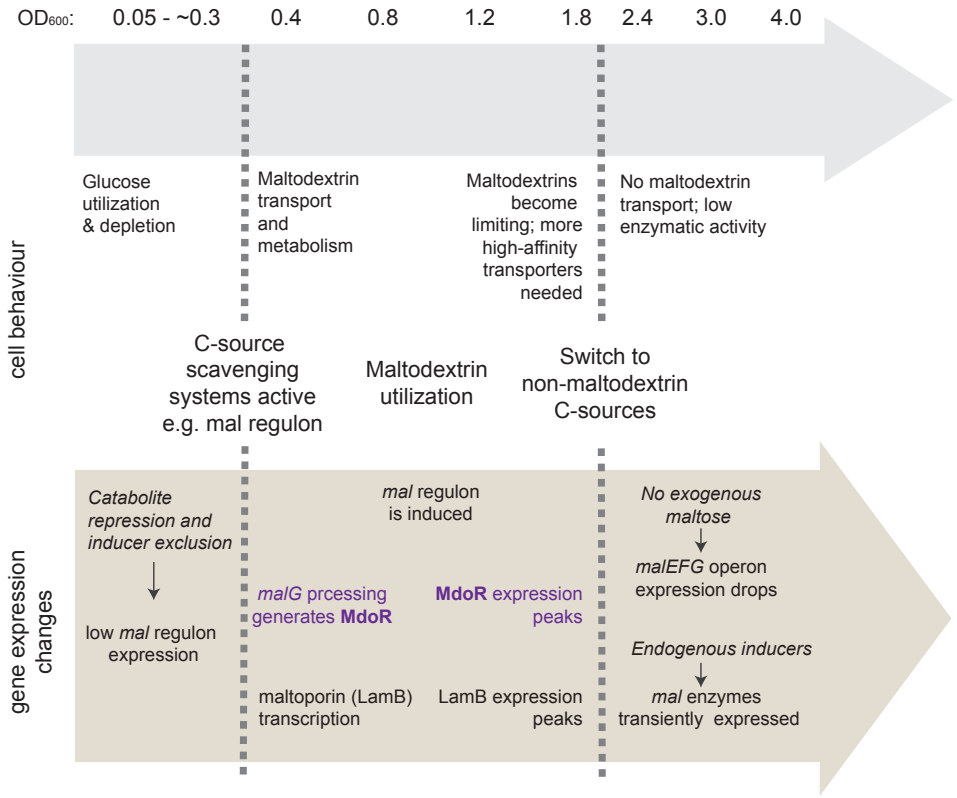

Fig. 4.11: A timeline of maltodextrin utilization in *E.coli* grown in LB. The cell behaviour and underlying gene regulation (bottom, green) correlated with OD_{600} (top, grey); the dashed lines indicate transitions in nutrient sensing and utilization; the purple text indicates post-

of exogenously-sourced maltotriose inducer, mal regulon genes are no longer transcribed. I observed that only the enzyme-encoding part of the *mal* regulon is briefly induced around OD_{600} 3.0. This is likely caused by formation of endogenous MalT inducers as intermediates of alternative carbon source metabolism (Schlegel et al., 2002) - the fact that this induction was not coupled with an increase in *lamB* and *malEFG* levels provides clear evidence that there is no maltodextrin intake at that stage of growth.

It is becoming clear that MdoR, directly derived from the *malEFG* mRNA, also peaks in conditions of intense maltodextrin assimilation and starts disappearing after OD_{600} 1.8. MdoR is still detectable at very low levels at OD_{600} 2.4 (Fig. 3.22) likely because it is stabilised by Hfq. I did not find many MdoR chimeric reads in the OD_{600} 1.8 samples, suggesting that MdoR is largely inactive in conditions of no maltose intake. Shortly after, MdoR is likely degraded as Hfq re-localizes to other RNAs required for

adaptation to a new environmental/metabolic context. The above observations were summed to reconstruct a time-course of intracellular events occurring with respect to maltose utilization, maltodextrin-transporter synthesis, and MdoR emergence under our growth conditions (Fig. 4.11).

Now that I have attained a better grasp of the environmental cues that regulate MdoR emergence and disappearance, namely presence and utilization of maltodextrins, my next aim is to define its role in these conditions.

4.9. MdoR contributes to the complex regulation of σ^E and OM composition during growth

The data discussed so far led me to the hypothesis that MdoR expression during late exponential phase results in down-regulation of a number of porin transcripts and sRNAs such as MicA. To test this, I analysed the levels of MdoR targets in the RNA-seq timecourse data. In the case of MicA, a sudden drop in steady-state at OD_{600} 1.8 relative to OD_{600} 1.2 is noticeable (Fig.4.12 A), so there is a chance that this decrease, coincident with the time-point of maximal MdoR expression, is controlled by this sRNA. At OD_{600} 2.4 (coincident with the disappearance of MdoR), sustained MicA increase with cell density is resumed. This pattern of expression is not observed in the other σ^E-dependent sRNAs, RybB and MicL (Fig. 4.12 A, D), suggesting that MicA specifically is targeted and repressed at OD_{600} 1.8. However, this needs to be confirmed by a biological replicate of the RNA-seq experiments. I do not have yet a biological replicate, so it is not possible to discern whether the decrease is biologically relevant or noise.

The pattern of *ompC* expression is much more difficult to interpret, especially that both its mRNA stability and translation efficiency are subject to post-transcriptional regulation. I noticed a modest decrease in *ompC* RNA levels in my Northern blot analyses (Fig. 4.12 C and E), but the RNAseq data did not confirm this result. Further, OD_{600}-dependent OmpC protein levels are largely inconsistent with mRNA levels (not

unexpected), but also do not show a sustained, inversely correlated decrease with MdoR increase (Fig. 4.12 D). Noteworthy, the observed OmpC expression profile is also not correlated with the expression profile of RybB (same for OmpA and MicA - Fig. 4.12 A and B).

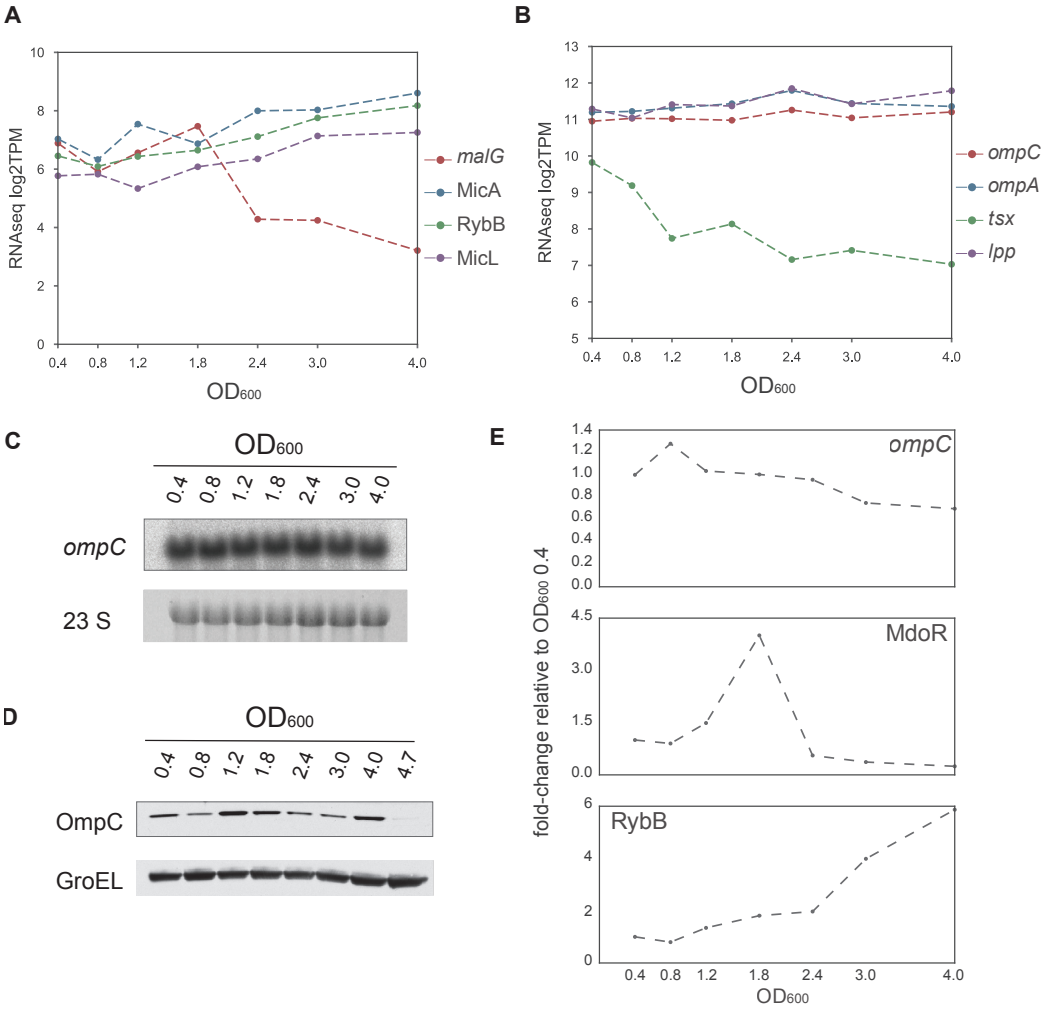

Fig. 4.12: Growth-stage dependent expression of MdoR targets, other OM proteins and σ^E regulon members in *E.coli*. (A) RNA steady-states (RNA-seq experiment) for σE regulon sRNAs and the MdoR precursor, *malG*. the y-axis shows normalized $\log_2$TPM (transcripts per million), the x-axis indicates the growth-stage in optical density units (OD_{600}). (B) RNA steady-states as in (A) of the messages encoding abundant OM components; (C) Northern blot analysis of *ompC* mRNA expression; OD_{600} is indicated above panels; 23S rRNA was used as the loading control. (D) Western blot analysis of OmpC levels; GroEL protein was used as the loading control. (E) Relative quantification of *ompC*, MdoR and RybB RNA using the OD_{600} 0.4 absolute quantification values as reference; the absolute quantification was performed with the AIDA software, using the Northern blot data in Fig. 3.22 and Fig. 4.12 C.

4.10. Discussion

The work described in this chapter revealed a number exciting prospects with respect to post-transcriptional regulation mechanisms and network circuitry involved in adaptive responses of *E.coli.*

The major discovery was the identification of MdoR, a novel riboregulator derived from the polycistronic *malEFG* precursor, that controls envelope composition and σ^E-mediated ESRs in response to nutrient availability. The characterization of MdoR adds a new node to the known, already complex network revolving around σ^E and OMPs (Fig. 4.1 versus Fig. 4.17).

MdoR biogenesis

It is unclear what factors trigger the first cleavage event, or at which site, but one possibility is the following: the *malEFG* operon is transcribed as a single primary transcript that is subsequently cleaved to uncouple its expression. This would allow differential post-transcriptional regulation of each product of the operon. This mode of regulation is common for ABC transporter operons, and is dictated by the gene order. Typically the gene product at the first position in the operon is made in much higher ratio relative to the other components (Belasco and Chen, 1988). In this case, MalE (encoded by *malE*), the periplasmic maltose-binding protein responsible high-affinity maltose transport has been reported to associate in multiple copies with the transmembrane complex (encoded by *malF* and *malG*). Indeed, I observed that even if the expression profile of all individual genes in the *malEFG* polycistronic transcript are similar (RNA-seq data in Fig 4.12A), *malE* is expressed at much higher levels (6-fold higher levels) than *malF* or *malG.* Henceforth, the degradation cascade might be initiated by the processing of *malEFG* produced from a single promoter, to yield a pool of independent transcripts: the distal gene *malE* may be clipped off and selectively stabilized (by protection from 3′-5′ exonucleases).

The rest of the transcript is more likely to be turned over, however, the 3'-UTR of *malG* is protected from degradation by Hfq. Indeed the selective stabilization of the *malE* cistron was previously reported, and is caused by the intercistronic stem loop found at the 3'-end of the stable *malE* (Newbury et al., 1987). It later was shown that maturation of the *malE* message requires both the activity of the degradosome (including RNase E), that cleaves the *malEFG* precursor and exonucleolytic cleavage of the 3'-stabilizing element (Khemici and Carpousis, 2004).

It is unclear whether the temperature shift causes MdoR to undergo targeted/regulated degradation or whether the degradation is due to increased RNase activity at higher temperatures. In both scenarios, the mechanism likely involves the displacement of Hfq from the Rho-independent terminator of MdoR to allow access for RNases. This is a reasonable speculation as, in conditions of heat-shock a few other stress-responsive sRNAs, such as RybB, are highly expressed. Hfq might relocate to prioritise the stabilisation and facilitate regulation by these sRNAs, thus rendering MdoR susceptible to RNases. In the absence of Hfq, MdoR is potentially very rapidly degraded due to its mature 5'P-end, a high-affinity substrate for RNase E (Mackie,

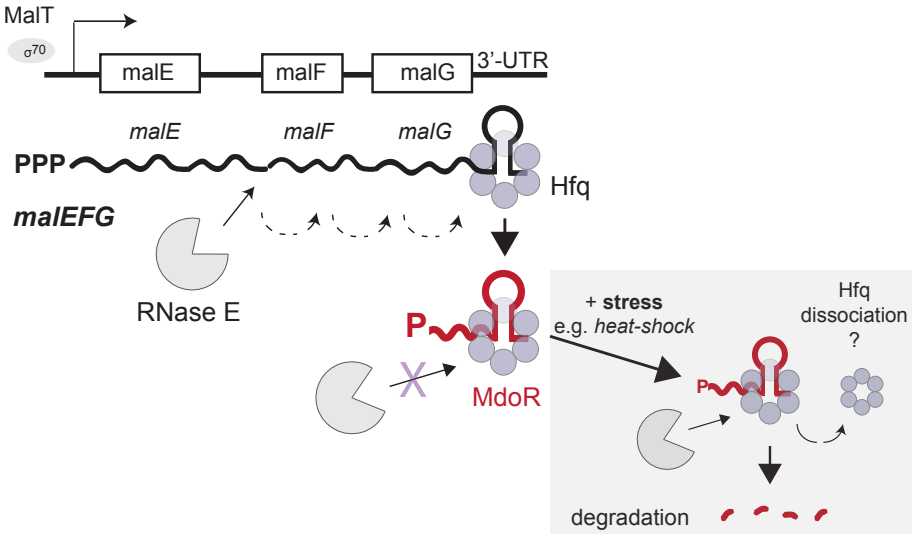

Fig. 4.13: A model for MdoR biogenesis from its *malEFG* mRNA precursor. The *malEFG* polycistron is initially cleaved in its upstream region by RNase E, which promotes subsequent rounds of degradation, that ultimately reach *malG*; its 3'-UTR (MdoR) is protected from degradation by Hfq binding at the Rho-independent terminator.'PPP' denotes triphosphate 5'-ends, whereas 'P' denotes monophosphate 5'-ends; the grey shaded area indicates a scenario in which RNase E is responsible for the destabilization of MdoR during stress (i.e. heat-shock), possibly involving Hfq dissociation from its terminator.

1998).

By summing up all the above observations, I formulate the following model for MdoR biogenesis and decay: a pool of the *malEFG* transcripts is constitutively degraded by RNase E, leaving behind 5'-P ends that promote further rounds of cleavage that affect *malG* (Fig. 4.5 A; Fig. 4.13). *MalG* 3'-UTRs binding Hfq are protected from RNase E-mediated decay, whereas in the absence of Hfq these fragments are susceptible to degradation. Upon imposition of stress, such as heat-shock, or potentially any other stress that requires a substantial rewiring of gene expression involving sRNA activity, Hfq dissociates from MdoR and re-locates to the stress-responsive sRNAs. Under these circumstances, MdoR is destabilised (Fig. 4.13; Fig. 4.5 C). Therefore, RNase E seems to be involved in both biogenesis and turnover of MdoR as part of two functionally distinct pathways by virtue of their temporal independence. The physiological cues that dictate the accumulation and disappearance of MdoR throughout/during standard growth were discussed in Section 4.8.

Approaches used for MdoR target identification

It is worth mentioning that during the characterization of MdoR, a combination of global analyses and validation approaches were employed. MdoR provides a case-study for a robust comparison of several sRNA targetome capture approaches (Fig. 4.8 C). Each methodology used for MdoR target prediction has its specifics, hence yields different datasets, but they did cross-validate each other for some genes. In retrospect, none of them failed to reveal OmpC as a target for MdoR - both i*n vivo* approaches (CLASH and pulse-overexpression studies) revealing it as the number one mRNA target of MdoR under our growth conditions. CLASH performed so well, that the potential base-pairing interactions of MdoR with its targets could immediately be accurately defined, which enhanced downstream seed mutation analyses.

Although not exhaustive, CLASH is the only approach that captured direct sRNA-mRNA interactions *in vivo*, and is the most informative with respect to seed sequence determination (see Chapter 3). Not only the CLASH data revealed both *ompC* and

ompA as targets, but also accurately defined the main seed sequence of MdoR - mutating the CLASH-predicted sequence of MdoR was sufficient to disrupt the regulatory function of this sRNA (Fig. 4.8).

CopraRNA uncovered that MdoR may be functional in other bacterial species (Section 4.5). I acknowledge that the methodology is prone to both false-positives and false-negatives and should be followed by experimental validations, but it is very informative with respect to the regulatory potential of a sRNA. Another limitation of this approach is that it will not report sRNA-sRNA interactions. Last, but not least, the computational prediction uncovers additional information about MdoR that is not determinable in our experimental conditions, including in non-model organisms. For example, many enzymes and other membrane-associated protein encoding mRNAs have been uncovered as putative MdoR targets. One reason these were not captured by either CLASH or overexpression studies may be that this study was limited to growth in LB.

The pulse-overexpression studies complement the CLASH data with information on the regulatory pathways that MdoR might be controlling/contributing to, a clear example of which is the σ^E network. only the pulse-overexpression experiment revealed a key player in our model, MicA as a MdoR target

Moreover, the CLASH data may favour long RNA-RNA base-pairing interactions as the enzymatic steps were performed at 37ºC, which may explain why we did not recover MicA-MdoR hybrids. Shorter bp interactions may be uncovered if we reduce the temperature for these steps.

This brings us to the point that there is no absolute approach to uncover the targetome of a sRNA. To unravel the role of MdoR to the extent thus far, I had to employ a suite of methods and extract the most relevant information to bring me to the proposed model. However, to complete the study, a few more key questions need to be answered (Section 4.12).

Possible regulatory effects of MdoR on its targets

The decrease in *ompC* RNA levels may either be the result of targeted RNase E-mediated decay, or destabilization because of translational inhibition. The exact mechanism remains to be elucidated and we are currently concentrating our efforts on performing *in vitro* binding and toeprinting assays to test whether MdoR directly interferes with *ompC* translation.

There are two alternative explanations for the incomplete shutoff of MdoR-mediated regulation of *ompC SM* (Section 4.7): I either mutated too many bases within the *ompC* sequence, which resulted in a change in the secondary structure that inhibits efficient MdoR regulation. Alternatively, both interacting regions within *ompC* mRNA may be required for full-regulatory effect of MdoR (only the more extensive region was mutated; Fig. 4.9 B).

Table 4.3: Genes with non-significant changes in expression upon MdoR overexpression in the DESeq2 data

Gene	log_2(Fold change)	DESeq2 adjusted p value
rybB	0.47	0.64
lamB	-0.24	0.72

Given that MicA is a σ^E-controlled sRNA, one route of indirect downregulation of MicA in the presence of MdoR would be via σ^E: MicA and RybB have the strongest σ^E promoters second only to the *rpoE* gene itself (Mutalik et al, 2009), so changes in σ^E levels could indirectly lead to changes in both MicA and RybB expression. Although *rpoE* mRNA levels were significantly decreased upon MdoR overexpression, if MicA down-regulation was indeed an indirect consequence of σ^E down-regulation, I would have expected that the fold-change in MicA levels would be less dramatic than that of *rpoE*, and perhaps similar to that of RybB (RybB was down-regulated in our dataset, but not at significant levels; Table 4.3). The differential expression data do not support this scenario. MicA levels were affected to a higher degree than could be explained by σ^E-mediated repression alone. It follows that MdoR either directly regulates MicA,

or indirectly downregulates it by other, unknown regulatory link, in addition to the σ^E-dependent pathways. The short time of MdoR induction supports the former hypothesis rather than the latter, and the confirmation of MdoR-MicA direct base-pairing is an immediate future aim of this study.

MdoR functionally links nutrient utilization to envelope-stress responses

I have established that MdoR is expressed at the transition between the exponential and stationary phases of growth in response to nutrient sensing. MdoR induction is dictated by the potency of exogenous maltodextrin assimilation in *E. coli*.

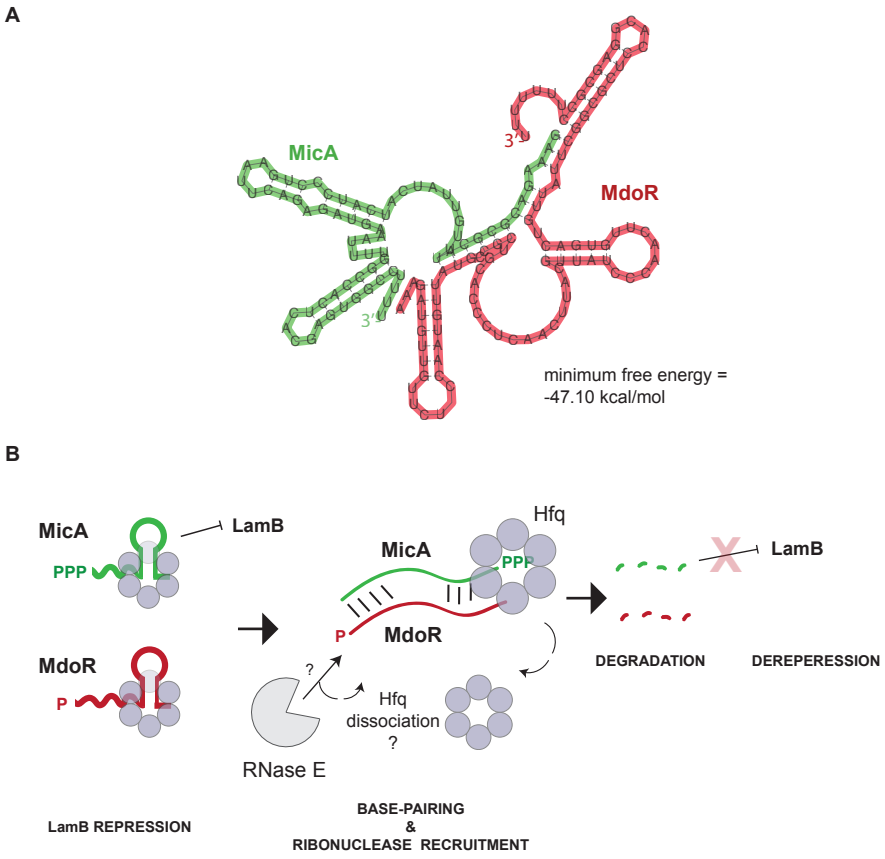

Fig. 4.14: A mechanism by which MdoR could mediate derepression of LamB. (A) A computationally predicted structure of the duplex between MdoR and MicA sRNAs (RNACofold, Lorenz et al., 2011); the minimum folding energy of the hybrid is indicated below the structure. (B) MicA and MdoR are both Hfq-stabilized trans-encoded sRNAs; as opposed to MicA, MdoR has a mature 5'-end (P); MicA directly downregulates LamB expression; MdoR base-pairing to MicA requires Hfq, but subsequent regulatory steps involve dissociation of Hfq from one or both sRNAs, either concomitantly or in a stepwise process; RNase E is recruited and the 5'-monophosphate terminus of MdoR promotes degradation of both MicA and MdoR, resulting in LamB derepression; '?' indicate that the mechanism or the precise involvement of regulators are not known.

I inferred that MdoR must have a function that would ultimately facilitate maltose uptake-related functions. This hypothesis motivated me to revisit the repertoire of direct and candidate targets of MdoR as revealed by CLASH and pulse-overexpression studies (Sections 3.11, 3.12 and 4.6). The most relevant candidate targets of MdoR that would support a role for MdoR during maltose utilization are MicA and *rpoE*, and the validated direct targets, *ompC* and *ompA*. These findings helped reconstruct a network that uncovers the rewiring of gene expression through two regulons, Mal and σ^E in response to nutrient sensing (Fig. 4.14 B). I will first discuss how the individual MdoR-components interaction contribute to the regulatory network, and move on to integrate them into a model.

As introduced in Fig. 4.2, MicA directly represses LamB expression in *E. coli* and thereby negatively regulates maltodextrin uptake (Bossi and Figueroa-Bossi, 2007). It follows that when cells feed on maltodextrin, MicA expression should be kept in check. MicA is a σ^E-dependent sRNA, whereas MdoR expression is not promoted by this σ factor, and is strikingly different from that of other σ^E-dependent sRNAs (Fig. 4.12 A, D). MicA was the second-most down-regulated target in the pulse-overexpression studies data and I hypothesize that MdoR directly base-pairs with MicA sRNA, resulting in sponging (Fig. 4.14 A, B). Although recently the sponging regulation by a 3'-UTR processed sRNAs was reported (Section 1.7; Myiakoshi et al., 2015b), a molecular understanding of the underlying mechanism is still lacking. It has been shown that RNase E and Hfq are both required, but the timing of the events was not described. It is not known whether both Hfq proteins corresponding to the interacting sRNAs are required for facilitating direct contacts, or only one Hfq protein remains associated with the regulating sRNA. At least one of the Hfq proteins has to be displaced to allow

degradation that is most likely mediated by RNase E assisted by the 5'-P of MdoR (Fig. 4.14 B). The regulatory outcome of the interaction is likely the degradation of both MicA and MdoR, resulting in LamB derepression.Even in absence of direct regulation, the DESeq2 data suggests that MdoR has such repressor potency on MicA (Fig. 4.9 B, C), to the extent that it could indirectly protect *lamB* from MicA-triggered degradation.

Indeed, I observed a minor (~20%), albeit not statistically significant, increase in *lamB* mRNA upon MdoR over-expression (Table 4.2). Given that sRNA overexpression was induced for only 15 minutes, the time window studied may not be enough to see dramatic changes in *lamB* levels. Towards this direction, I plan to induce MdoR over-expression over a longer time-course and quantify *lamB* expression.

I deem the MdoR-triggered repression of σ^E as relevant to the physiological context of membrane composition changes in accordance to metabolic decisions, in this case, the boost in maltodextrin assimilation. Any perturbations in membrane composition are sensed and trigger σ^E activation, which directly results in stress-responsive high production of *rpoE* and MicA, the two genes with the strongest σ^E promoters (Mutalik et al, 2009). By repressing σ^E, MdoR may not only negatively regulate MicA specifically, but also indirectly prevent its accumulation. Therefore, MdoR may dampen/alleviate a σ^E-mediated response that would otherwise lower LamB synthesis. This additional fine-tuning of σ^E expression becomes even more important if MdoR indeed acts as a sponge on MicA - as this regulation mode would most likely result in concomitant turnover of both MicA and MdoR, leading to removal of a fraction of the regulator.

The σ^E repression observed during MdoR overexpression is likely an indirect effect of reduced OMP synthesis (OmpC, OmpA, Tsx). OmpC is a major non-specific porin that allows the passage of multiple substrates, including maltose - albeit at significantly lower diffusion rates than LamB. This makes OmpC a poor choice of maltodextrin transporter when the external concentration drops below a certain threshold. During maltodextrin assimilation, there is a gradual switch between expression of non-specific maltose porins (low rate of maltose diffusion) to the expression of the 100-fold more efficient maltose-specific porin, LamB (Szmelcman and Hofnung, 1975; Szmelcman et al., 1976). Since MdoR expression peaks at OD_{600} 1.8, just before the exhaustion of LB maltodextrin content (concomitant with the peak in *lamB* expression; Fig. 4.12 A; Fig. 4.13), I propose that it directly opposes OmpC synthesis as a response to the increased need for maltose-specific transport systems. At the same time, *ompC, ompA* and *tsx* are very abundant transcripts and the synthesis and assembly of their

products require an overwhelming fraction of a cell's chaperones and assembly factors (Section 4.1). Consequently, the reduction in OMP synthesis mediated by MdoR would render more of these factors available for the correct folding and integration into the membrane for maltose-specific transport system components, such as LamB. By triggering these changes in outer membrane protein expression, MdoR activity itself may contribute to the membrane modifications that could induce the aforementioned σ^E activation. However, as already mentioned, MdoR mitigates the ESR induced by σ^E activity by repressing *rpoE* expression.

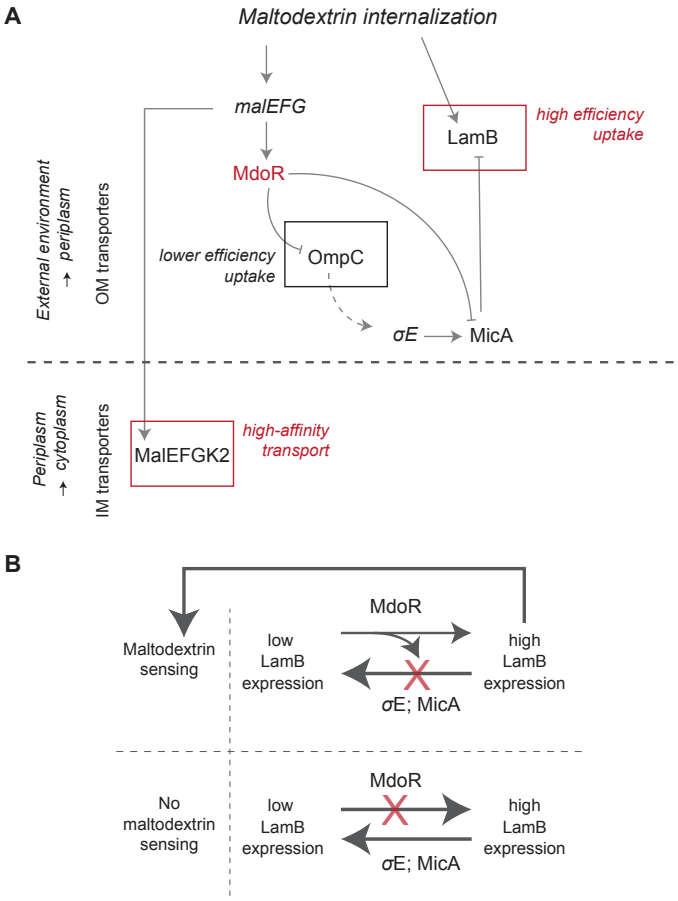

Fig. 4.15: MdoR mediates the switch from lower-efficiency maltose uptake to high efficiency uptake of maltodextrins into the periplasmic space by bridging a connection between C-source utilization and stress-responses in *E.coli.* (A) Proposed model of the MdoR-mediated regulatory network that leads to the derepression of LamB (via MicA and σ^E); the activation of both coding and non-coding arms encoded at the malEFG locus results *malEFG and lamB* mRNAs cross-talk and coordinated expression of IM, OM and periplasmic components of the maltose-specific transport system. 'IM' and 'OM' denote the inner membrane and outer membrane localization of the regulated transporters (boxes), respectively; the text in italics indicates which step of maltodextrin transport they mediate.; the dashed line indicates homeostatic control (B) Physiological consequences of LamB post-transcriptional regulation; during maltodextrin utilization (top), MdoR antagonizes MicA to increase LamB synthesis, thereby enhancing maltodextrin sensing (feed-forward loop); when maltodextrins are not assimilated (bottom), the MicA-mediated repression of LamB is active, resulting in low LamB expression.

Noteworthy, the MdoR-mediated repression of σ^E via OMP synthesis repression appears as the single-most effective way of increasing LamB synthesis and facilitate its correct assembly into the OM. OmpC and OmpA repression is necessary to decrease competition for folding factors in favour of the newly synthesized LamB, whose expression is indirectly promoted by MdoR by removal of MicA repression. In principle, cells could tackle the increased need for envelope assembly complexes by boosting σ^E activation and expression to produce more of these factors. However, activating σ^E at an inappropriate time can have deleterious effects. Furthermore, in this particular case, σ^E activation could defeat the purpose of MdoR-mediated downregulation of MicA.

The proposed model is that MdoR is a non-coding component of the maltose-specific transport system of *E.coli,* that functionally helps shift the balance between synthesis of lower-efficiency and higher-efficiency maltodextrin uptake systems in favor of that of maltose-specific porins, while dampening σ^E controlled ESRs (Fig. 4.15 A).
Strictly with respect to maltose utilization, MdoR activity has two important consequences that ultimately result in the net more efficient maltose assimilation. First, MdoR contributes to the coordinated expression of IM, and OM-components of the maltose transport system (Fig. 4.15 A). During maltose utilization at very low exogenous concentrations, it may be unproductive to downregulate the OM maltoporin (via MicA) while increasing IM transporters levels (the coding arm of *malEFG*). This is reasonable assumption, as the two have co-dependent functions and MalE, the periplasmic maltose-binding protein responsible for affinity has physical links to both (Shuman, 1982). In this scenario, MicA repression by MdoR would allow the crosstalk between two physically unlinked genes, but that are physiologically linked. Secondly, MdoR contributes to the amplification of the signal of the MalT feed-forward loop (Fig. 4.10 B): more efficient maltodextrin intake due to elevated LamB levels leads to the production of more MalT inducers, thereby activating transcription of the entire Mal regulon. Overall, it becomes clear that all extracytoplasmic components of the maltose transport system are not only genetically and physically linked, but also connected post-transcriptionally. In conclusion, MdoR is a small RNA that regulates membrane

composition and ESRs in response to nutrient availability sensing.

Physiological MdoR induction effect on the levels of its targets

The results in Section 4.9 motivates me to question how much these sRNAs actually contribute to porin regulation *in vivo*; clearly, the endogenous levels of these sRNAs are not sufficient to completely shut down OMP synthesis under our growth conditions (Fig. 4. 12). However, even a slight decrease in OmpC and OmpA translation may result in a phenotypically-relevant effect due to their high stability and abundance, as was proposed for MicL sRNA effect on the synthesis of Lpp, the most abundant lipoprotein (Guo et al., 2014). In line with this rationale, even a small reduction in OMP synthesis caused by MdoR expression may relieve enough pressure on the OMP

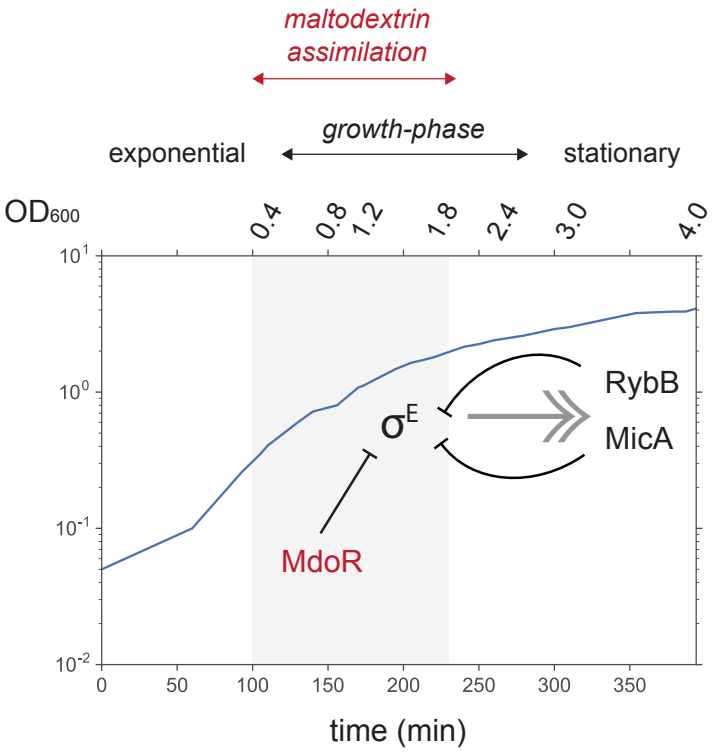

Fig. 4.16: A model that proposes a OD_{600}-dependent control of σ^E by OMP-regulating sRNAs in *E.coli*. The growth-curve data is taken from the CLASH experiment, and the sampled OD_{600} and corresponding growth-stage assignment are indicated above the plot; σ^E is known to be repressed by the homeostatic negative-feedback loop rendered by the σ^E-transcribed (grey arrow) sRNAs, MicA and RybB, a feedback loop more predominant at later stages of growth (because σ^E is induced upon entry into stationary phase). MdoR is a sRNA that is expressed at earlier stages of growth that also represses σ^E, but not as part of a negative feedback loop - since MdoR is not produced by σ^E; MdoR expression is instead induced during maltodextrin utilization, which under the growth conditions

assembly machinery to re-direct folding factors to LamB (e.g. SecA, Sec complex, SurA, Skp, YaeI etc; Ureta et al., 2007). Moreover, the mere concomitant presence of an sRNA and its target inside the cell at a given point does not automatically imply the two RNAs are interacting at that particular point - as pointed earlier, it is very difficult to find the growth conditions at which a regulator and its target interact. Finally, there may be other, unidentified sRNAs and other RBPs that regulate OMP expression (Chapter 3). Given the complexity of porin regulation and the large number of transcriptional and post-transcriptional regulators involved (Section 4.1, Fig. 4.1), I cannot point to a single sRNA as the sole responsible regulator for the observed OmpC levels during growth. Nevertheless, I intend to gain more insights into the mechanism by which MdoR represses OmpC synthesis (Sections 4. 6; Section 4.7)

Throughout growth, *rpoE* and *rseA* displayed a similar expression pattern as the σ^E-dependent sRNA MicL, and most likely MdoR activity alone cannot explain the observed profile (Fig.4.12). As it is the case for OMPs, σ^E expression may be regulated by other post-transcriptional regulators, some of them unknown (Chapter 3). Indeed, it was previously shown that cells lacking Hfq dramatically induce the σ^E-mediated ESR (Figueroa-Bossi et al., 2006, Guisbert et al, 2007) - and I believe this observation is not only an effect of destabilization of MicA and RybB as it was initially proposed. To our knowledge, MdoR is the first sRNA shown to repress σ^E, that is not produced by σ^E itself. Downregulation of σ^E expression (Fig. 4.16) is wired via the homeostatic negative loop caused by MdoR-mediated repression of the OMP synthesis - this already indicates a temporally-defined 'division of labour' among the sRNAs that control *rpoE* expression. It was initially apparent that this division was OD_{600}-dependent: MdoR could repress OMPs and σ^E at earlier stages of growth, whereas at later stages of growth σ^E was mainly regulated by negative feedback from RybB and MicA, also via OMP repression - a model consistent with the expression profiles of these sRNAs (Fig 4.12 A, D). It was not known, however, what environmental factor accounts for σ^E repression by each of the two mechanisms. It later became clear that maltodextrin availability sensing was the main MdoR-inducing factor (Section 4.8).

In conditions of severe stress (e.g. heat-shock) that require activation of σ^E for cell-survival, MdoR is degraded by a mechanism that may involve re-localization of Hfq and RNAse E-mediated decay (Section 4.4). Thus, the MdoR inhibitory effect on σ^E is relieved, allowing increased σ^E expression that allows expression of stress-responsive genes. Under these conditions, σ^E expression is fine-tuned via a negative feedback-loop by the small RNAs that are part of its own regulon: MicA, RybB and MicL.

A model for MdoR-mediated regulation

MdoR is placed in a strategic position at the interface between the MalT and σ^E regulons, that connects nutrient sensing and stress adaptation in *E. coli* (Fig. 4.17). These findings uncover insights into how σ^E network is regulated in response to nutrient availability, an exciting avenue that was not explored as extensively as the σ^E-

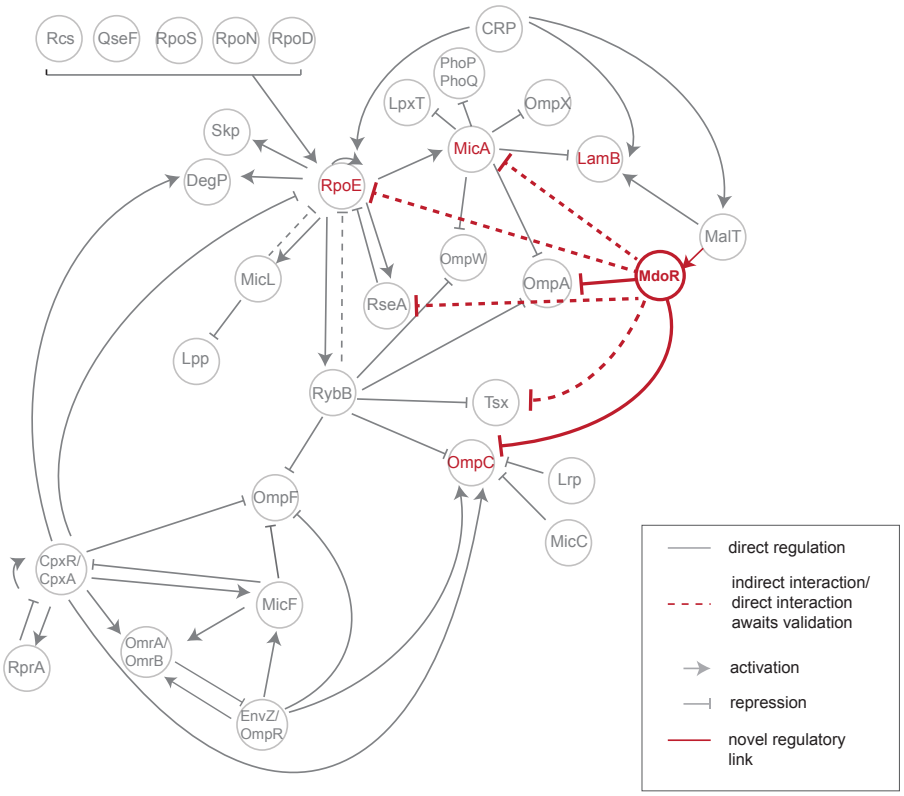

Fig. 4.17: An updated network of σ^E and OMP-regulation, that includes MdoR as a new node. Novel connections are shown in red, whereas known interactions (also shown in Fig. 4.1) are shaded in grey.

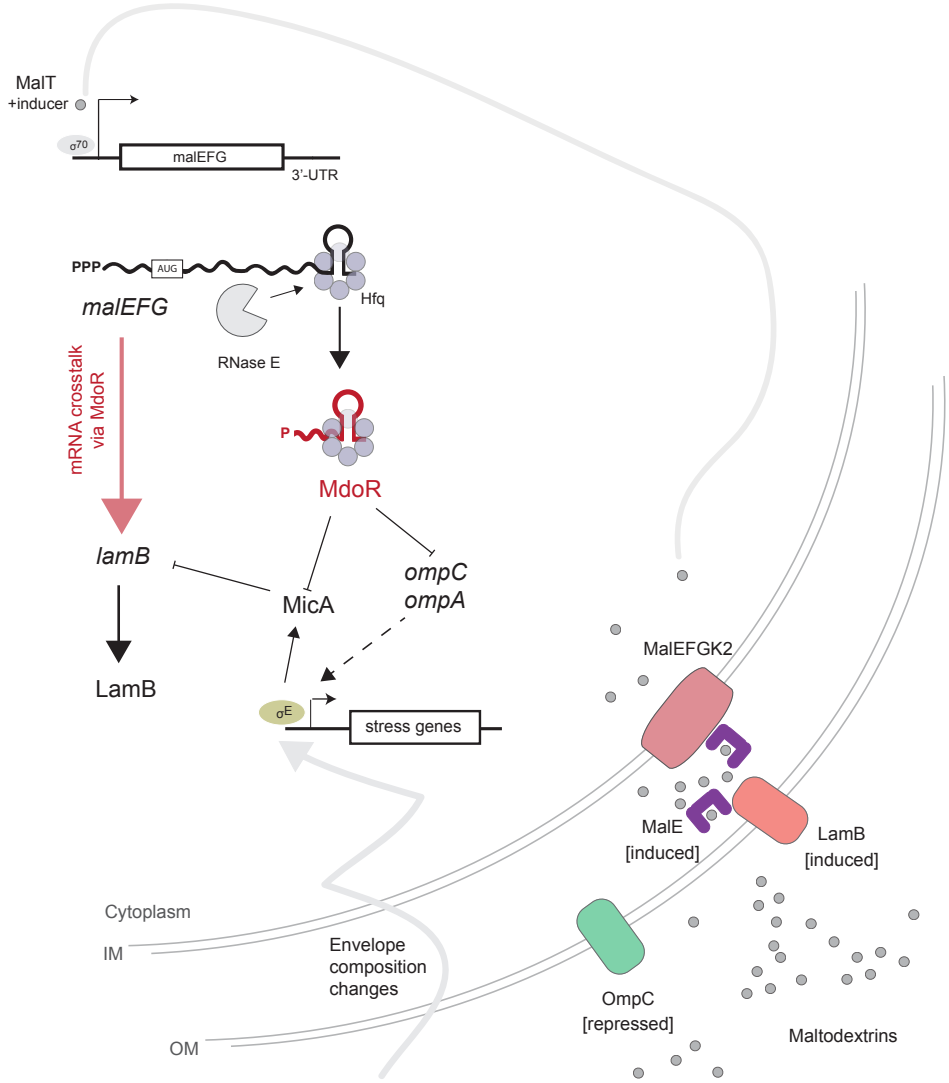

Fig. 4.18: A model that summarizes MdoR biogenesis and regulatory role. During maltodextrin utilization, the *malEFG* operon, encoding IM components for maltose uptake is transcribed by the housekeeping sigma factor (σ^{70}) and MalT-inducer. The coding arm of the *malEFG* locus leads to the expression of high-affinity maltose ABC transporters and maltose-binding protein (purple), which transfer maltodextrins via the IM. A fraction of the *mallEFG* transcript pool is constitutively degraded by RNase E, generating MdoR, a regulatory sRNA derived from the 3'-UTR of *malG*. MdoR is stabilised by the RNA chaperone Hfq that protects it from further ribonucleolytic cleavage. MdoR is the non-coding arm of a pathway that contributes to increased efficiency of maltodextrin uptake by a mechanism that involves the fine-tuning of a stress response with respect to C-source utilization. MdoR base-pairs with *ompC* and *ompA* mRNAs, leading to their decay. This results in repression of OmpC, a porin with limited ability to translocate maltodextrins across the OM. Any changes in outer membrane composition are sensed and trigger the activation of the envelope-stress alternative sigma factor (σ^E). However, the activation of σ^E leads to an increase in MicA levels, a small RNA that strongly downregulates LamB, a maltoporin that is very efficient in maltose uptake. MdoR derepresses LamB synthesis by destabilizing MicA, and by indirect downregulation of σ^E (via OmpC and OmpA). The downregulation of porins may increase the availability of assembly factors for insertion of newly-synhesised LamB into the membrane. Overall, MdoR mediates the crosstalk between *malEFG* and *lamB* mRNAs (red arrow). The net outcome of mal regulon transcription, MdoR biogenesis and regulatory activity, is increased expression of high-affinity components of maltose-specific transport (MalE and LamB). This results in higher diffusion rates of maltodextrins inside the cell.

monitored maintenance of envelope integrity during stress. MdoR specifically responds to maltodextrin sensing, and its expression peaks with increased necessity for high-affinity maltodextrin transporters when this carbon source becomes limited, but is still available (Section 4.8). The activity of MdoR results in decreased expression of general, non-specific porins (OmpC, OmpA) and derepression of LamB (maltose-specific porin) synthesis. Thus, physiologically, the activation of MdoR results in improved utilization and sensing of maltodextrin substrates (Section 4.9). These regulatory events are linked to σ^E repression, and that maintains the directionality/ consistency of MdoR-mediated regulation and posits a functional link between the Mal and σ^E regulons (Fig. 4.18).

Moreover, these regulatory outcomes ensue from mechanisms with a degree of novelty. Although the repression of OmpC and OmpA is most likely caused by canonical, Hfq-facilitated base-pairing to their 5'-UTR resulting in transcript destabilization (Section 4.7), LamB derepression is indirect and ensues from a more atypical mode of regulation. MicA sRNA is known to target *lamB* mRNA to reduce LamB synthesis. MdoR had a remarkably strong (and immediate) repressing effect on MicA expression (Section 4.6), and I propose that MdoR directly base-pairs with MicA, leading to its destabilization and subsequent MicA-targetome (including LamB) derepression (Section 4.9). If that is the case, LamB net upregulation would be a consequence of sponging, by a mechanism similar to that described for SroC (Miyakoshi et al., 2015) and *chb* mRNA (Figueroa-Bossi et al., 2009). Indeed, the 5'-end of MdoR bears a monophosphate that would facilitate RNase E recruitment and degradation of both sRNAs. Thus, MdoR mediates communication between the physically unlinked, yet functionally linked, *malEFG* and and *lamB* mRNAs, by controlling MicA turnover. It is not known at this point whether the link between σ^E regulon, MicA and MdoR is bidirectional, namely, if either σ^E or MicA can, in specific conditions regulate *malEFG*/ MdoR expression. It is striking, however, that during heat-shock, a strong σ^E-inducing condition, MdoR was degraded (Section 4.4) - this would result in σ^E derepression. This possibility should be further explored, as it would uncover two different regulatory inputs that regulate the balance between the two sRNAs (regulator becomes the

regulated, possibly MicA or another component of the MicA regulon could directly or indirectly degrade MdoR).

Interestingly, according to the model, RNase E is likely involved in a total of three distinct events revolving around MdoR. First, RNase E processes a polycistronic parental RNA to produce the active form of MdoR. The mechanistic details of MdoR biogenesis place this sRNA in the recently-growing list of Type II 3'-UTR derived sRNAs that are stabilized by Hfq (Chao et al., 2012). Second, RNase E seems to be involved in the turnover of MdoR during stress conditions (Section 4.4). Third, RNase E is the main candidate ribonuclease that would facilitate MdoR-sponging of MicA.

I have uncovered a previously-unsuspected physiological connection between maltodextrin utilization (controlled by MdoR) and envelope stress (via MicA and σ^E). MdoR - mediated regulation combines a plethora of network circuitry modes: direct repression of general porins and homeostatic feedback to σ^E, crosstalk between *malEFG* and *lamB*, and coherent feed-forward looping to MalT. Moreover, by its repression of MicA, MdoR has the potential to regulate the entire MicA regulon, an avenue I have not explored yet. I speculate however, that the effect of MdoR on other targets of MicA may be correlated with its ability to compete with those targets for MicA base-pairing, ability fine-tuned by MdoR production rate, relative expression levels but also thermodynamic requirements for binding.

Noteworthy, MdoR possesses features that were only recently described and considered at the time "non-canonical" - this vision needs to be updated as more riboregulators with such features are being described. If its sponge activity is proven, MdoR would be another 3'-UTR-derived sRNA that acts as a sponge - and I speculate there is a correlation between the biogenesis mode of sRNAs and their mode of regulation.

The work on MdoR characterization advances the field not only by its discovery as a novel riboregulator, but contributes to our understanding of RNA-based circuits and

integration of input signals, as exemplified in this work by the integration of σ^E and Mal regulon responses.

Future plans

The immediate future directions that will follow up this thesis will focus on accumulating experimental evidence to support the proposed model (Fig. 4.19) and to gain more mechanistic insights into MdoR-mediated regulation.

First, I plan to induce overexpression of MdoR for a longer time-course and monitor changes in *lamB* mRNA - this experiment would provide indirect evidence that MdoR is responsible for net LamB upregulation, possibly by MicA sponging. To prove the latter, the direct base-pairing of MdoR with MicA has to be validated by *in vitro* binding assays, in the presence of Hfq or alternatively the co-expression of both sRNAs from plasmids combined with mutational analyses of the interacting regions. These studies could be followed by demonstration of Hfq and RNase E-dependency of the sponging *in vivo.*

To further dissect the direct repression of OmpC by MdoR, we are currently employing EMSA, which could be followed by both secondary structure probing and toeprinting experiments. Additionally, I will test the requirement of Hfq for the interaction in Hfq knock-out strains. In order to obtain a more comprehensive understanding of MdoR effect on the translation of multiple protein-coding targets, together with our collaborator, Gabriella Viero we are currently performing polysome-profiling analyses in MdoR pulse-overexpression conditions, which will be followed up by quantitative RT-PCR of individual targets. Last, but not least, experiments that confirm MdoR expression during maltodextrin sensing/utilization conditions, time-course shifts from no-maltodextrin containing media to maltodextrin-containing media followed by Northern Blot/RT-qPCR will reveal the pattern of MdoR expression in more defined growth conditions. Anti-MalE Western blotting will be indicative of the intensity of maltose utilization at each time-point.

5. Concluding remarks

Recent years have witnessed outstanding progress in our understanding of how post-transcriptional regulation works: state of the art techniques and pipelines for data analyses have been developed and launched us into a revolution of RNA networks interpretation.

Although we (think we) have a good idea which and how many sRNAs these networks comprise, to unravel their function, a good representation of their targets is needed. Most of our understanding of sRNA networks derived from characterization of individual sRNAs and subsets of their targets. Recently, by adopting methodologies for direct mapping of RNA-RNA interactions *in vivo* from eukaryotes, comprehensive networks of sRNA-target RNA interactions were generated in *E.coli* (Section 1.8). The intermolecular RNA interactions identified in association with Hfq and RNase E significantly expanded the complexity of sRNA networks (Fig. 1.9). However, these data reflect interactions for a defined set of growth conditions, so the interactions uncovered are not universal. For example, three growth conditions were investigated by RIL-seq: exponential growth phase, stationary growth phase and low-iron conditions, but only a subset of the uncovered interactions were discovered in all growth conditions (Melamed et al., 2016). Moreover, using RNase E as a bait to capture RNA-RNA interactions reveals only the sets of RNA-RNA pairs where RNase E is required for regulation, and this method was applied only in virulence-inducing conditions, in a pathogenic *E.coli* strain (Waters et al., 2017)

Summary of the main findings

In this work, I presented new data that uncovers the dynamics of Hfq binding to sRNAs and mRNAs, and cumulates the intermolecular RNA-RNA interactions occurring during standard growth in *E.coli*. I investigated a time-course that reflects the transitions between the exponential and stationary phases of growth.

In Chapter 3, I performed a set of bioinformatic analyses that highlighted growth stage-dependent trends in Hfq binding to RNAs, and a collection of several intriguing examples of Hfq-mediated regulation. Notably, these findings had little overlap with the RIL-seq data, probably caused by both the differences in growth conditions studied, and the experimental steps. The validation of most of my findings is needed, but the predictions and speculations I made about them present exciting prospects worth investigating.

The global analyses I present are in good agreement with published reports, with respect to both the Hfq-recognition sites, and motifs found within chimeras.

However, compared to other studies, my experimental design as a time-course offered me the opportunity to study changes in Hfq binding from a kinetic perspective, and to compare these with the changes in total RNA levels. This analysis was the starting point from which I selected potentially interesting examples of atypical sRNAs. Some of these case-studies suggest that *cis*-acting sRNAs could be also *trans*-acting sRNAs, and even act as sponges that may control the activity of other sRNA (Section 3.8). Moreover, an sRNA conventionally known to sponge proteins - CsrB, might also act as a sponge for sRNAs. Both examples, if proven correct, would illustrate how cells exploit the versatility of (single) RNA regulators to 'multi-task'. I propose this strategy would allow cells to coordinate multiple responses, increase the robustness of a response, or prevent the overshoot of an adaptive response - all with a low metabolic cost.

The analysis of a model sRNA, ChiX, revealed that this sRNA, detectable at all growth-stages, can be involved in multiple sRNA-sRNA interactions (Section 3. 7). Although it cannot be concluded from the CLASH data which sRNA acts as the sponge within the interactions, conjectures can be made. In the illustrated example (Fig. 3.14), I propose that another sRNA, ArrS, may trap ChiX in a duplex, to prevent ChiX-mediated regulation of a mRNA targets, but not another. This is a base-pairing pattern that suggests an additional parameter that dictates which targets from an sRNA regulon

are affected by the sponging activity, besides the relative expression of the interacting species and the thermodynamic details of the interaction. Further investigation of the sRNA seeds occluded by such interactions, and analysis of their equivalent on cognate mRNAs, would help delineate the regulatory potential of sRNA sponges. Compared to the RIL-seq data, Hfq CLASH data uncovered many interactions between sRNAs. This finding is intriguing because it implicates direct cross-talk between regulons of independently-transcribed sRNAs. As opposed to most described sponges in the literature, which are processed sRNAs, the interactions between sRNAs that are primary transcripts would allow a different type of post-transcriptional tuning, more likely to be characterised by reversibility. One example I investigated more closely is the interaction between MgrR and GcvB - it suggests that the two sRNA pathways, known to be indirectly linked (Section 3.8) - may also be directly linked, possibly with GcvB forming an incoherent feedforward loop to regulate MgrR expression. It would be interesting to see how many cases of this type of regulation there really are.

Mechanistically, both CLASH and RIL-seq uncovered that 3'-UTRs of mRNAs are a common source of sRNAs - however, the parental mRNA species predicted, and the intermolecular interactions identified are very different (Section 3.9). Some of these sRNAs may require RNase E to be released from their precursors (Chao et al., 2017). One such transcript is the 3'-UTR of *malG.* The hybrids uncovered by CLASH suggested that this RNA fragment might be involved in the regulation of major porins in *E. coli* (OmpC, OmpA). The expression profile of this sRNA is very different from that of other porin-regulating sRNAs, RybB and MicA (Fig. 3.22), and is distinguishable by a spike in its level at the transition between exponential and stationary phases of growth. To unravel the role of this sRNA and to understand why it base-pairs with porin mRNAs at this particular growth stage, I characterised it in detail.

Chapter 4 dealt with the experimental characterization of the novel malG 3'-UTR-derived sRNA, which I named MdoR. MdoR is generated by processing from the last gene of a maltose operon that encodes the inner-membrane maltose transporter. I validated experimentally the *in vivo* interaction of MdoR with *ompC* 5'-UTR. The seed

sequence predicted by CLASH streamlined the validation, as it precisely mapped where MdoR base-pairs with these targets. The expression pattern of MdoR already suggested that this sRNA is not redundant with RybB and MicA. Consequently, to unravel its complete set of targets and to uncover what pathways MdoR is part of, I performed pulse-overexpression studies (Section 4.5). The short time-scale of these experiments strongly suggests that the differentially expressed genes are direct MdoR targets. Besides reconfirming *ompC* and *ompA* as targets, I observed a very strong downregulation of MicA upon MdoR overexpression. MicA is an envelope-stress responsive sRNA known to repress expression of lamB - encoding the outer-membrane porin for maltodextrin uptake. I propose that MdoR could act on MicA to derepress LamB. Corroborating this finding with the research I made on understanding when maltodextrins are utilized throughout growth (Section 4.7), helped me develop the following model: to enhance maltodextrin utilization, MdoR represses the non-coding arm of an envelope stress response that otherwise suppresses the expression of the maltodextrin porin. At the same time, MdoR reduces the expression of general porins, possibly to increase availability of porin assembly factors. Physiologically, this results in more efficient maltodextrin uptake - as MdoR fine-tunes a stress-response pathway in response to nutrient availability. Moreover, MdoR illustrates how characterization of a piece of data from a CLASH dataset can lead to a better understanding of the pathways that build the regulatory networks in bacteria.

Broader implications for the field

The application of both CLASH and RIL-seq in bacteria have the potential to revolutionize how we view the rewiring of gene expression through regulatory networks.

They are powerful tools that consistently hint at the idea that mechanisms previously considered non-canonical, are in fact, common. Not only is the bacterial genome much richer in regulators, but also, even the known regulators could employ more diverse roles than envisioned. Mechanisms that were described for only one RNA class, could be performed by another class. The latter is to a certain degree intuitive,

as they are all the same type of macromolecule, with unifying features - RNAs of any kind, by design, are prone to interact with both RBPs and other nucleic acids. It does make sense that microorganisms would exploit this property - even more so when conditions are not optimal (and the resources limited), such as during stress. The efficiency of the regulatory strategies is also reflected by how even degradation products are used as regulatory molecules.

This study also brings forward the importance of having accompanying datasets when performing transcriptome-wide studies. For instance, employing RNA-seq in parallel with CLASH can reveal insights into how well the binding of a protein to RNA is correlated with their steady states, but could also provide a snapshot of the expression levels of RNAs involved in an interaction. The latter would augment predictive models of the outcome of RNA-RNA base-pairing. Alternatively, performing Hfq CLASH in parallel with RNase E CLASH (or other RBPs) could help unravel the dynamics and the involvement of these proteins in RNA-mediated regulation. Last, but not least, RNA-seq data can be used as a readout of the state the cells are in - crucial for integrating any findings into the broader physiological context of the studied organism.

A next challenge for the field is to integrate these novel RNA networks with those of protein regulators and study correlations with proteomics data. This is currently not possible as RNA-RNA interaction mapping methods cannot directly detect the directionality or the effect (activating or repressing) of the regulation. Predictions can prove helpful in this direction - but ultimately, the gold-standard for understanding an interaction is its experimental characterization. However, these large datasets are an astounding resource for finding interesting/prioritising which interactions to investigate (as I attempted to illustrate in Chapter 3).

REFERENCES

Acuña, L. G., Barros, M. J., Peñaloza, D., Rodas, P. I., Paredes-Sabja, D., Fuentes, J. A., … Calderón, I. L. (2016). A feed-forward loop between SroC and MgrR small RNAs modulates the expression of eptB and the susceptibility to polymyxin B in Salmonella Typhimurium. Microbiology (United Kingdom), 162(11), 1996–2004. https://doi.org/10.1099/mic.0.000365

Ades, S. E., Connolly, L. E., Alba, B. M., & Gross, C. A. (1999). The Escherichia coli sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. Genes Dev, 13(18), 2449–2461. https://doi.org/10.1111/j.1574-695X.1996.tb00291.x

Aiba, H. (2007). Mechanism of RNA silencing by Hfq-binding small RNAs. Current Opinion in Microbiology, 10(2), 134–139. https://doi.org/10.1016/j.mib.2007.03.010

Aiso, T., Kamiya, S., Yonezawa, H., & Gamou, S. (2014). Overexpression of an antisense RNA, ArrS, increases the acid resistance of Escherichia coli. Microbiology (Reading, England), 160(2014), 954–961. https://doi.org/10.1099/mic.0.075994-0

Alba, B. M., & Gross, C. A. (2004). Regulation of the Escherichia coli sigmaE-dependent envelope stress response. Molecular Microbiology, 52(3), 613–619. https://doi.org/10.1111/j.1365-2958.2003.03982.x

Alon, U. (2007). Network motifs: theory and experimental approaches. Nature Reviews Genetics, 8(6), 450–461. https://doi.org/10.1038/nrg2102

Altuvia, S., Zhang, A., Argaman, L., Tiwari, A., & Storz, G. (1998). The Escherichia coli OxyS regulatory RNA represses fhlA translation by blocking ribosome binding. EMBO Journal, 17(20), 6069–6075. https://doi.org/10.1093/emboj/17.20.6069

Amar, A., Mamun, M. Al, Lombardo, M., Shee, C., Andreas, M., Gonzalez, C., … Susan, M. (2012). Identity and Function of a Large Gene Network Underlying Mutagenic Repair of DNA Breaks. Science, 338(6112), 1344–1348. https://doi.org/10.1126/science.1226683. Identity

Argaman, L., Argaman, L., Hershberg, R., Hershberg, R., Vogel, J., Vogel, J., … Altuvia, S. (2001). Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. Curr Biol, 11(12), 941–950. https://doi.org/10.1016/j.gene.2008.09.024

Arraiano, C. M., Andrade, J. M., Domingues, S., Guinote, I. B., Malecki, M., Matos, R. G., … Viegas, S. C. (2010). The critical role of RNA processing and degradation in the control of gene expression. FEMS Microbiology Reviews, 34(5), 883–923. https://doi.org/10.1111/j.1574-6976.2010.00242.x

Azam, M. S., & Vanderpool, C. K. (2015). RNA Sponges Mediate Cross Talk Between Functionally Related Messenger RNAs. EMBO J 34(11), 1436–1438. https://doi.org/10.15252/embj.201591492

Baev, M. V., Baev, D., Jansco Radek, A., & Campbell, J. W. (2006). Growth of Escherichia coli MG1655 on LB medium: Monitoring utilization of amino acids, peptides, and nucleotides with transcriptional microarrays. Applied Microbiology and Biotechnology, 71(3), 317–322. https://doi.org/10.1007/s00253-005-0310-5

Baker, C. S., Eöry, L. A., Yakhnin, H., Mercante, J., Romeo, T., & Babitzke, P. (2007). CsrA inhibits translation initiation of Escherichia coli hfq by binding to a single site overlapping the Shine-Dalgarno sequence. Journal of Bacteriology, 189(15), 5472–5481. https://doi.org/10.1128/JB.00529-07

Balbontín, R., Fiorini, F., Figueroa-Bossi, N., Casadesús, J., & Bossi, L. (2010). Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in salmonella enterica. Molecular Microbiology, 78(2), 380–394. https://doi.org/10.1111/j.1365-2958.2010.07342.x

Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y. I., … Collado-Vides, J. (2009). Regulation by transcription factors in bacteria: Beyond description. FEMS Microbiology Reviews, 33(1), 133–151. https://doi.org/10.1111/j.1574-6976.2008.00145.x

Bandyra, K. J., Bouvier, M., Carpousis, A. J., & Luisi, B. F. (2013). The social fabric of the RNA degradosome. Biochimica et Biophysica Acta - Gene Regulatory Mechanisms, 1829(6–7), 514–522. https://doi.org/10.1016/j.bbagrm.2013.02.011

Bandyra, K. J., Said, N., Pfeiffer, V., Górna, M. W., Vogel, J., & Luisi, B. F. (2012). The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. Molecular Cell, 47(6), 943–953. https://doi.org/10.1016/j.molcel.2012.07.015

Bartel, D. P. (2009). MicroRNA Target Recognition and Regulatory Functions. Cell, 136(2), 215–233. https://doi.org/10.1016/j.cell.2009.01.002.MicroRNA

Beisel, C. L., & Storz, G. (2011). The Base-Pairing RNA Spot 42 Participates in a Multioutput Feedforward Loop to Help Enact Catabolite Repression in Escherichia coli. Molecular Cell, 41(3), 286–297. https://doi.org/10.1016/j.molcel.2010.12.027

Benjamin, J. A. M., & Mass??, E. (2014). The iron-sensing aconitase B binds its own mRNA to prevent sRNA-induced mRNA cleavage. Nucleic Acids Research, 42(15), 10023–10036. https://doi.org/10.1093/nar/gku649

Bilusic, I., Popitsch, N., Rescheneder, P., Schroeder, R., & Lybecker, M. (2014). genome through Hfq co-immunoprecipitation, (May), 641–654.

Bobrovskyy, M., & Vanderpool, C. K. (2016). Diverse mechanisms of post-transcriptional repression by the small RNA regulator of glucose-phosphate stress. Molecular Microbiology, 99(2), 254–273. https://doi.org/10.1111/mmi.13230

Bohn, C., Rigoulay, C., & Bouloc, P. (2007). No detectable effect of RNA-binding protein Hfq absence in Staphylococcus aureus. BMC Microbiology, 7, 10. https://doi.org/10.1186/1471-2180-7-10

Boos, W., & Shuman, H. (1998). Maltose / Maltodextrin System of Escherichia coli : Transport , Metabolism , and Regulation Maltose / Maltodextrin System of Escherichia coli : Transport , Metabolism , and Regulation. Microbiology and Molecular Biology Reviews, 62(1), 204–229.

Bossi, L., & Figueroa-Bossi, N. (2007). A small RNA downregulates LamB maltoporin in Salmonella. Molecular Microbiology, 65(3), 799–810. https://doi.org/10.1111/j.1365-2958.2007.05829.x

Bossi, L., & Figueroa-Bossi, N. (2016). Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. Nature Reviews Microbiology, 14(12):775-784. doi: 10.1038/nrmicro.2016.129

Bossi, L., Schwartz, A., Guillemardet, B., Boudvillain, M., & Figueroa-Bossi, N. (2012). A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. Genes and Development, 26(16), 1864–1873. https://doi.org/10.1101/gad.195412.112

Bouvier, M., Sharma, C. M., Mika, F., Nierhaus, K. H., & Vogel, J. (2008). Small RNA Binding to 5′ mRNA Coding Region Inhibits Translational Initiation. Molecular Cell, 32(6), 827–837. https://doi.org/10.1016/j.molcel.2008.10.027

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. Current Opinion in Microbiology, 10(2), 102–109. https://doi.org/10.1016/j.mib.2007.03.012

Brennan, R. G., & Link, T. M. (2007). Hfq structure, function and ligand binding. Current Opinion in Microbiology, 10(2), 125–133. https://doi.org/10.1016/j.mib.2007.03.015

Byun, Y., & Han, K. (2009). PseudoViewer3: Generating planar drawings of large-scale RNA structures with pseudoknots. Bioinformatics, 25(11), 1435–1437. https://doi.org/10.1093/bioinformatics/btp252

Carpousis, A. J. (2007). The RNA Degradosome of Escherichia coli : An mRNA-Degrading Machine Assembled on RNase E. Annual Review of Microbiology, 61(1), 71–87. https://doi.org/10.1146/annurev.micro.61.080706.093440

Cavanagh, A. T., Chandrangsu, P., & Wassarman, K. M. (2010). 6S RNA regulation of relA alters ppGpp levels in early stationary phase. Microbiology, 156(12), 3791–3800. https://doi.org/10.1099/mic.0.043992-0

Chaba, R., Alba, B. M., Guo, M. S., Sohn, J., Ahuja, N., Sauer, R. T., & Gross, C. a. (2011). Signal integration by DegS and RseB governs the σE-mediated envelope stress response in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 108(5), 2106–2111. https://doi.org/10.1073/pnas.1019277108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1019277108

Chae, H., Han, K., Kim, K. S., Park, H., Lee, J., & Lee, Y. (2011). Rho-dependent termination of ssrS (6S RNA) transcription in Escherichia coli: Implication for 3' processing of 6S RNA and expression of downstream ygfA (putative 5-formyl-tetrahydrofolate cyclo-ligase). Journal of Biological Chemistry, 286(1), 114–122. https://doi.org/10.1074/jbc.M110.150201

Chao, Y. (2016). Article A 3'UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Article A 3 0 UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response, 1–12. https://doi.org/10.1016/j.molcel.2015.12.023

Chao, Y., & Vogel, J. (2010). The role of Hfq in bacterial pathogens. Current Opinion in Microbiology, 13(1), 24–33. https://doi.org/10.1016/j.mib.2010.01.001

Chao, Y., Li, L., Girodat, D., Forstner, K. U., Said, N., Corcoran, C., … Vogel, J. (2017). in vivo cleavage map illuminates the central cole of RNase E in coding and non-coding RNA pathways. Molecular Cell, 65(1), 39–51. https://doi.org/10.1016/j.molcel.2016.11.002

Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C. M., & Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. The EMBO Journal, 31(20), 4005–4019. https://doi.org/10.1038/emboj.2012.229

Chen, H., Bjerknes, M., Kumar, R., & Jay, E. (1994). Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of E. coli mRNAs. Nucleic Acids Research, 22(23), 4953–4957. https://doi.org/10.1093/nar/22.23.4953

Chen, S., Zhang, A., Blyn, L. B., & Storz, G. (2004). MicC , a Second Small-RNA Regulator of Omp Protein Expression in Escherichia coli MicC , a Second Small-RNA Regulator of Omp Protein Expression in Escherichia coli. Journal of Bacteriology, 186(20), 6689–6697. https://doi.org/10.1128/JB.186.20.6689

Church, G. M., Church, G. M., Rosenow, C., & Rosenow, C. (2003). Global RNA Half-Life Analysis in. Genome Research, (January), 216–223. https://doi.org/10.1101/gr.912603.

Clarke, J. E., Kime, L., Romero A., D., & McDowall, K. J. (2014). Direct entry by RNase e is a major pathway for the degradation and processing of RNA in Escherichia coli. Nucleic Acids Research, 42(18), 11733–11751. https://doi.org/10.1093/nar/gku808

Coornaert, A., Chiaruttini, C., Springer, M., & Guillier, M. (2013). Post-Transcriptional Control of the Escherichia coli PhoQ-PhoP Two-Component System by Multiple sRNAs Involves a Novel Pairing Region of GcvB. PLoS Genetics, 9(1), 12–14. https://doi.org/10.1371/journal.pgen.1003156

Coornaert, A., Lu, A., Mandin, P., Springer, M., Gottesman, S., & Guillier, M. (2010). MicA sRNA links the PhoP regulon to cell envelope stress. Molecular Microbiology, 76(2), 467–479. https://doi.org/10.1111/j.1365-2958.2010.07115.x

Corcoran, C. P., Podkaminski, D., Papenfort, K., Urban, J. H., Hinton, J. C. D., & Vogel, J. (2012). Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. Molecular Microbiology, 84(3), 428–445. https://doi.org/10.1111/j.1365-2958.2012.08031.x

Davis, B. M., Quinones, M., Pratt, J., Ding, Y., & Waldor, M. K. (2005). Characterization of the small untranslated RNA RyhB and its regulon in Vibrio cholerae. Journal of Bacteriology, 187(12), 4005–4014. https://doi.org/10.1128/JB.187.12.4005-4014.2005

De La Cruz, M. Á., & Calva, E. (2010). The complexities of porin genetic regulation. Journal of Molecular Microbiology and Biotechnology, 18(1), 24–36. https://doi.org/10.1159/000274309

De Las Peñas, A., Connolly, L., & Gross, C. A. (1997). The sigmaE-mediated response to extracytoplasmic stress in Escherichia coli is transduced by RseA and RseB, two negative regulators of sigmaE. Molecular Microbiology, 24(2), 373–385. https://doi.org/9159523

De Lay, N., & Gottesman, S. (2009). The crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. Journal of Bacteriology, 191(2), 461–476. https://doi.org/10.1128/JB.01157-08

De Lay, N., Schu, D. J., & Gottesman, S. (2013). Bacterial small RNA-based negative regulation: Hfq and its accomplices. Journal of Biological Chemistry, 288(12), 7996–8003. https://doi.org/10.1074/jbc.R112.441386

Desnoyers, G., Bouchard, M., & Masse, E. (2012). New insights into small RNA- dependent translational regulation in prokaryotes, 29(2), 1–7. https://doi.org/10.1016/j.tig.2012.10.004

Desnoyers, G., Morissette, A., Prévost, K., & Massé, E. (2009). Small RNA-induced

differential degradation of the polycistronic mRNA iscRSUA. The EMBO Journal, 28(11), 1551–1561. https://doi.org/10.1038/emboj.2009.116

Deutscher, M. P. (2006). Degradation of RNA in bacteria: Comparison of mRNA and stable RNA. Nucleic Acids Research, 34(2), 659–666. https://doi.org/10.1093/nar/gkj472

Dippel, R., & Boos, W. (2005). The Maltodextrin System of Escherichia coli : Metabolism and Transport T. Journal of Bacteriology, 187(24), 8322–8331. https://doi.org/10.1128/JB.187.24.8322

Dressaire, C., Picard, F., Redon, E., Loubière, P., Queinnec, I., Girbal, L., & Cocaign-Bousquet, M. (2013). Role of mRNA Stability during Bacterial Adaptation. PLoS ONE, 8(3). https://doi.org/10.1371/journal.pone.0059059

Dressaire, C., Picard, F., Redon, E., Loubière, P., Queinnec, I., Girbal, L., & Cocaign-Bousquet, M. (2013). Role of mRNA Stability during Bacterial Adaptation. PLoS ONE, 8(3). https://doi.org/10.1371/journal.pone.0059059

Dubey, A. K., Baker, C. S., Romeo, T., & Babitzke, P. (2005). CsrA − RNA interaction RNA sequence and secondary structure participate in high-affinity CsrA – RNA interaction. Rna, (11), 1579–1587. https://doi.org/10.1261/rna.2990205.3

Duval, M., Simonetti, A., Caldelari, I., & Marzi, S. (2015). Multiple ways to regulate translation initiation in bacteria: Mechanisms, regulatory circuits, dynamics. Biochimie, 114, 18–29. https://doi.org/10.1016/j.biochi.2015.03.007

Elf, J., Hampel, K., Zimmermann, B., & Wagner, E. G. H. (2010). RNAs actively cycle on the Sm-like protein Hfq, Genes Develop (2010).pdf, (Table 1), 2621–2626. https://doi.org/10.1101/gad.591310.Structures

Ellis, M. J., Trussler, R. S., & Haniford, D. B. (2015). Hfq binds directly to the ribosome-binding site of IS10 transposase mRNA to inhibit translation. Molecular Microbiology, 96(3), 633–650. https://doi.org/10.1111/mmi.12961

Emory, S. A., & Belasco, J. G. (1990). The ompA 5' untranslated RNA segment functions in Escherichia coli as a growth-rate-regulated mRNA stabilizer whose activity is unrelated to translational efficiency. Journal of Bacteriology, 172(8), 4472–4481.

Faucher, S. P., & Shuman, H. A. (2011). Small regulatory RNA and Legionella pneumophila. Frontiers in Microbiology, 2(MAY), 1–14. https://doi.org/10.3389/fmicb.2011.00098

Fei, J., Singh, D., Zhang, Q., Park, S., Balasubramanian, D., Vanderpool, C. K., … Biology, C. (2015). HHS Public Access, 347(6228), 1371–1374. https://doi.org/10.1126/science.1258849.Determination

Ferenci, T. (1996). Adaptation to life at micromolar nutrient levels: The regulation of Escherichia coli glucose transport by endoinduction and cAMP. FEMS Microbiology Reviews, 18(4), 301–317. https://doi.org/10.1016/0168-6445(96)00019-8

Ferrario, M., Ernsting, B. R., Borst, D. W., Wiese, D. E., Blumenthal, R. M., & Matthews, R. G. (1995). The leucine-responsive regulatory protein of Escherichia colinegatively regulates transcription ofompCandmicFand positively regulates translation ofompF. J Bacteriol, 177(1), 103–113.

Figueroa-Bossi, N., Lemire, S., Maloriol, D., Balbotin, R., Casadesus,J. & Bossi, L. (2006). Loss of Hfq activates the σE-dependent envelope stress response in Salmonella

enterica. Mol Microbiology 62(3), 838-852

Figueroa-Bossi, N., Valentini, M., Malleret, L., & Bossi, L. (2009). Caught at its own game: Regulatory small RNA inactivated by an inducible transcript mimicking its target. Genes and Development, 23(17), 2004–2015. https://doi.org/10.1101/gad.541609

Flint, A., Butcher, J., & Stintzi, A. (n.d.). Stress Responses, Adaptation, and Virulence of Bacterial Pathogens During Host Gastrointestinal Colonization. Virulence Mechanisms of Bacterial Pathogens, Fifth Edition, 387–411. https://doi.org/10.1128/microbiolspec.VMBF-0007-2015

Fozo, E. M., Kawano, M., Fontaine, F., Kaya, Y., Kathy, S., Jones, K. L., … Storz, G. (2008). NIH Public Access, 70(5), 1076–1093. https://doi.org/10.1111/j.1365-2958.2008.06394.x.Repression

Fröhlich, K. S., & Vogel, J. (2009). Activation of gene expression by small RNA. Current Opinion in Microbiology, 12(6), 674–682. https://doi.org/10.1016/j.mib.2009.09.009

Fröhlich, K. S., Papenfort, K., Fekete, A., & Vogel, J. (2013). A small RNA activates CFA synthase by isoform-specific mRNA stabilization. The EMBO Journal, 32(22), 2963–2979. https://doi.org/10.1038/emboj.2013.222

Gangaraju Vamsi K. Lin Haifan. (2009). NIH Public Access. Nat Rev Mol Cell Biol., 10 (2)(1), 116–125. https://doi.org/10.1038/nrm2621

García-Martínez, J., Aranda, A., & Pérez-Ortín, J. E. (2004). Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. Molecular Cell, 15(2), 303–313. https://doi.org/10.1016/j.molcel.2004.06.004

García-Martínez, J., Ayala, G., Pelechano, V., Chávez, S., Herrero, E., & Pérez-Ortín, J. E. (2012). The relative importance of transcription rate, cryptic transcription and mRNA stability on shaping stress responses in yeast. Transcription, 3(1), 39–44. https://doi.org/10.4161/trns.3.1.19416

Geiger, T., Francois, P., Liebeke, M., Fraunholz, M., Goerke, C., Krismer, B., … Wolz, C. (2012). The Stringent Response of Staphylococcus aureus and Its Impact on Survival after Phagocytosis through the Induction of Intracellular PSMs Expression. PLoS Pathogens, 8(11). https://doi.org/10.1371/journal.ppat.1003016

Geiger, T., Francois, P., Liebeke, M., Fraunholz, M., Goerke, C., Krismer, B., … Wolz, C. (2012). The Stringent Response of Staphylococcus aureus and Its Impact on Survival after Phagocytosis through the Induction of Intracellular PSMs Expression. PLoS Pathogens, 8(11). https://doi.org/10.1371/journal.ppat.1003016

Georg, J., & Hess, W. R. (2011). cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. Microbiology and Molecular Biology Reviews, 75(2), 286–300. https://doi.org/10.1128/MMBR.00032-10

Gogol, E. B., Rhodius, V. A., Papenfort, K., Vogel, J., & Gross, C. A. (2011). Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. Proceedings of the National Academy of Sciences, 108(31), 12875–12880. https://doi.org/10.1073/pnas.1109379108

González-Flecha, B., & Demple, B. (1999). Role for the oxyS Gene in Regulation of Intracellular Hydrogen Peroxide in E. coli. J. Bacteriol, 181(12), 3833–3836. Retrieved from http://jb.asm.org/content/181/12/3833%5Cnhttp://jb.asm.org/content/181/12/3833#ref-list-1

Görke, B., & Vogel, J. (2008). Noncoding RNA control of the making and breaking of sugars Noncoding RNA control of the making and breaking of sugars, 2914–2925. https://doi.org/10.1101/gad.1717808

Gottesman, S. (2005). Micros for microbes: Non-coding regulatory RNAs in bacteria. Trends in Genetics, 21(7), 399–404. https://doi.org/10.1016/j.tig.2005.05.008

Gottesman, S., & Storz, G. (2011). Bacterial Small RNA Regulators : Versatile Roles and Rapidly Evolving Variations, 1–16. https://doi.org/10.1101/cshperspect.a003798

Grabowicz, M., & Silhavy, T. J. (2017). Envelope Stress Responses: An Interconnected Safety Net. Trends in Biochemical Sciences, 42(3), 232–242. https://doi.org/10.1016/j.tibs.2016.10.002

Granneman, S., Kudla, G., Petfalski, E., & Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proceedings of the National Academy of Sciences, 106(24), 9613–9618. https://doi.org/10.1073/pnas.0901997106

Guillier, M., & Gottesman, S. (2008). The 5′ end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. Nucleic Acids Research, 36(21), 6781–6794. https://doi.org/10.1093/nar/gkn742

Guillier, M., Gottesman, S., Storz, G., Guillier, M., Gottesman, S., & Storz, G. (2006). Modulating the outer membrane with small RNAs Modulating the outer membrane with small RNAs, 2338–2348. https://doi.org/10.1101/gad.1457506

Guisbert, E., Rhodius, V. A., Ahuja, N., Witkin, E., & Gross, C. A. (2007). Hfq modulates the σE-mediated envelope stress response and the σ32-mediated cytoplasmic stress response in Escherichia coli. Journal of Bacteriology, 189(5), 1963–1973. https://doi.org/10.1128/JB.01243-06

Guo, M. S., Updegrove, T. B., Gogol, E. B., Shabalina, S. A., Gross, C. A., & Storz, G. (2014). MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. Genes and Development, 28(14), 1620–1634. https://doi.org/10.1101/gad.243485.114

Han, K., Tjaden, B., & Lory, S. (2016). GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. Nature Microbiology, 2(December 2016), 16239. https://doi.org/10.1038/nmicrobiol.2016.239

Hansen, M. J., Chen, L. H., Fejzo, M. L. S., & Belasco, J. G. (1994). the OmpA 5' Untranslated Region Impedes a Major Pathway for Messenger-Rna Degradation in Escherichia-Coli. Molecular Microbiology, 12(5), 707–716. https://doi.org/10.1111/j.1365-2958.1994.tb01058.x

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., & Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature, 495(7441), 384–388. https://doi.org/10.1038/nature11993

Hayes, E. T., Wilks, J. C., Sanfilippo, P., Yohannes, E., Tate, D. P., Jones, B. D., … Slonczewski, J. L. (2006). Oxygen limitation modulates pH regulation of catabolism and hydrogenases, multidrug transporters, and envelope composition in Escherichia coli K-12. BMC Microbiology, 6, 89. https://doi.org/10.1186/1471-2180-6-89

Helwak, A., & Tollervey, D. (2014). Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). Nature Protocols, 9(3), 711–728. https://doi.org/10.1038/nprot.2014.043

Helwak, A., Kudla, G., Dudnakova, T., & Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell, 153(3), 654–665. https://doi.org/10.1016/j.cell.2013.03.043

Henkin, T. M. (2008). Riboswitch RNAs : using RNA to sense cellular metabolism Riboswitch RNAs : using RNA to sense cellular metabolism, 3383–3390. https://doi.org/10.1101/gad.1747308

Holmqvist, E., & Vogel, J. (2013). A small RNA serving both the Hfq and CsrA regulons. Genes and Development, 27(10), 1073–1078. https://doi.org/10.1101/gad.220178.113

Holmqvist, E., Reimegård, J., Sterk, M., Grantcharova, N., Römling, U., & Wagner, E. G. H. (2010). Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. The EMBO Journal, 29(11), 1840–1850. https://doi.org/10.1038/emboj.2010.73

Holmqvist, E., Wright, P. R., Li, L., Bischler, T., Barquist, L., Reinhardt, R., … Vogel, J. (2016). Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. The EMBO Journal, 35(9), e201593360. https://doi.org/10.15252/embj.201593360

Holt, L. J., Lyons, R. J., Ryan, A. S., Beale, S. M., Ward, A., Cooney, G. J., & Daly, R. J. (2009). Dual ablation of Grb10 and Grb14 in mice reveals their combined role in regulation of insulin signaling and glucose homeostasis. Molecular Endocrinology (Baltimore, Md.), 23(9), 1406–14. https://doi.org/10.1210/me.2008-0386

Hör, J., & Vogel, J. (2017). Global snapshots of bacterial RNA networks. The EMBO Journal, 36(3), 245–247. https://doi.org/10.15252/embj.201696072

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Hussein, R., & Lim, H. N. (2011). Disruption of small RNA signaling caused by competition for Hfq. Proceedings of the National Academy of Sciences, 108(3), 1110–1115. https://doi.org/10.1073/pnas.1010082108

Hussein, R., & Lim, H. N. (2012). Direct comparison of small RNA and transcription factor signaling. Nucleic Acids Research, 40(15), 7269–7279. https://doi.org/10.1093/nar/gks439

Hwang, W., Arluison, V., & Hohng, S. (2011). Dynamic competition of DsrA and rpoS fragments for the proximal binding site of Hfq as a means for efficient annealing. Nucleic Acids Research, 39(12), 5131–5139. https://doi.org/10.1093/nar/gkr075

Ikeda, Y., Yagi, M., Morita, T., & Aiba, H. (2011). Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli. Molecular Microbiology, 79(2), 419–432. https://doi.org/10.1111/j.1365-2958.2010.07454.x

Ishihama, A. (2010). Prokaryotic genome regulation: Multifactor promoters, multitarget regulators and hierarchic networks. FEMS Microbiology Reviews, 34(5), 628–645. https://doi.org/10.1111/j.1574-6976.2010.00227.x

Ishihama, A., Kori, A., Koshio, E., Yamada, K., Maeda, H., Shimada, T., … Fujitac, N. (2014). Intracellular concentrations of 65 species of transcription factors with known regulatory functions in Escherichia coli. Journal of Bacteriology, 196(15), 2718–2727. https://doi.org/10.1128/JB.01579-14

Jäger, D., Pernitzsch, S. R., Richter, A. S., Backofen, R., Sharma, C. M., & Schmitz, R. A. (2012). An archaeal sRNA targeting cis-and trans-encoded mRNAs via two distinct domains. Nucleic Acids Research, 40(21), 10964–10979. https://doi.org/10.1093/nar/gks847

Jagodnik, J., Chiaruttini, C., & Guillier, M. (2017). Stem-Loop Structures within mRNA Coding Sequences Activate Translation Initiation and Mediate Control by Small Regulatory RNAs. Molecular Cell, 68(1), 158–170.e3. https://doi.org/10.1016/j.molcel.2017.08.015

Jiang, X., & Belasco, J. G. (2004). Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. Proceedings of the National Academy of Sciences of the United States of America, 101(25), 9211–9216. https://doi.org/10.1073/pnas.0401382101

Johansen, J., Rasmussen, A. A., Overgaard, M., & Valentin-Hansen, P. (2006). Conserved Small Non-coding RNAs that belong to the ??E Regulon: Role in Down-regulation of Outer Membrane Proteins. Journal of Molecular Biology, 364(1), 1–8. https://doi.org/10.1016/j.jmb.2006.09.004

Jørgensen, M. G., Thomason, M. K., Havelund, J., Jørgensen, M. G., Havelund, J., … Storz, G. (2013). Dual function of the McaS small RNA in controlling biofilm formation Dual function of the McaS small RNA in controlling biofilm formation, 1132–1145. https://doi.org/10.1101/gad.214734.113

Jost, D., Nowojewski, A., & Levine, E. (2013). Regulating the Many to Benefit the Few: Role of Weak Small RNA Targets. Biophysical Journal, 104(8), 1773–1782. https://doi.org/10.1016/J.BPJ.2013.02.020

Joyce, S. a, & Dreyfus, M. (1998). In the absence of translation, RNase E can bypass 5' mRNA stabilizers in Escherichia coli. Journal of Molecular Biology, 282(2), 241–254. https://doi.org/10.1006/jmbi.1998.2027

Kalamorz, F., Reichenbach, B., März, W., Rak, B., & Görke, B. (2007). Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli. Molecular Microbiology, 65(6), 1518–1533. https://doi.org/10.1111/j.1365-2958.2007.05888.x

Kasner, E., Hunter, C. A., Ph, D., Kariko, K., & Ph, D. (2013). NIH Public Access, 70(4), 646–656. https://doi.org/10.1002/ana.22528.Toll-like

Kawano, M., Reynolds, A. A., Miranda-Rios, J., & Storz, G. (2005). Detection of 5′- and 3′-UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli. Nucleic Acids Research, 33(3), 1040–1050. https://doi.org/10.1093/nar/gki256

Khemici, V., & Carpousis, A. J. (2004). The RNA degradosome and poly(A) polymerase of Escherichia coli are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers. Molecular Microbiology, 51(3), 777–790. https://doi.org/10.1046/j.1365-2958.2003.03862.x

Kime, L., Clarke, J. E., Romero A., D., Grasby, J. A., & McDowall, K. J. (2014). Adjacent single-stranded regions mediate processing of tRNA precursors by RNase e direct entry. Nucleic Acids Research, 42(7), 4577–4589. https://doi.org/10.1093/nar/gkt1403

Klein, G., & Raina, S. (2017). Small regulatory bacterial RNAs regulating the envelope stress response. Biochemical Society Transactions, 45(2), 417–425. https://doi.org/10.1042/BST20160367

Kortmann, J., & Narberhaus, F. (2012). Bacterial RNA thermometers: molecular zippers and switches. Nature Reviews Microbiology, 10(4), 255–265. https://doi.org/10.1038/nrmicro2730

Kudla, G., Granneman, S., Hahn, D., Beggs, J. D., & Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. Proceedings of the National Academy of Sciences, 108(24), 10010–10015. https://doi.org/10.1073/pnas.1017386108

Lalaouna, D., Carrier, M. C., Semsey, S., Brouard, J. S., Wang, J., Wade, J. T., & Massé, E. (2015). A 3' external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. Molecular Cell, 58(3), 393–405. https://doi.org/10.1016/j.molcel.2015.03.013

Laubacher, M. E., & Ades, S. E. (2008). The Rcs phosphorelay is a cell envelope stress response activated by peptidoglycan stress and contributes to intrinsic antibiotic resistance. Journal of Bacteriology, 190(6), 2065–2074. https://doi.org/10.1128/JB.01740-07

Lee, T., & Feig, A. L. (2008). The RNA binding protein Hfq interacts specifically with tRNAs The RNA binding protein Hfq interacts specifically with tRNAs, 514–523. https://doi.org/10.1261/rna.531408.Hfq

Levine, E., & Hwa, T. (2008). Small RNAs establish gene expression thresholds. Current Opinion in Microbiology, 11(6), 574–579. https://doi.org/10.1016/j.mib.2008.09.016

Levine, E., Zhang, Z., Kuhlman, T., & Hwa, T. (2007). Quantitative characteristics of gene regulation by small RNA. PLoS Biology, 5(9), 1998–2010. https://doi.org/10.1371/journal.pbio.0050229

Li, Z., & Deutscher, M. P. (2002). RNase E plays an essential role in the maturation of Escherichia coli tRNA precursors. RNA (New York, N.Y.), 8(1), 97–109. https://doi.org/10.1017/S1355838202014929

Li, Z., Pandit, S., & Deutscher, M. P. (1999). RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. EMBO Journal, 18(10), 2878–2885. https://doi.org/10.1093/emboj/18.10.2878

Ling, J., Cho, C., Guo, L. T., Aerni, H. R., Rinehart, J., & Söll, D. (2012). Protein Aggregation Caused by Aminoglycoside Action Is Prevented by a Hydrogen Peroxide Scavenger. Molecular Cell, 48(5), 713–722. https://doi.org/10.1016/j.molcel.2012.10.001

Lisitsky, I., & Schuster, G. (1999). Preferential degradation of polyadenylated and polyuridinylated RNAs by the bacterial exoribonuclease polynucleotide phosphorylase. European Journal of Biochemistry, 261(2), 468–474. https://doi.org/10.1046/j.1432-1327.1999.00285.x

Liu, D., Chang, X., Liu, Z., Chen, L., & Wang, R. (2011). Bistability and oscillations in gene regulation mediated by small noncoding RNAs. PLoS ONE, 6(3). https://doi.org/10.1371/journal.pone.0017029

Liu, M. Y., & Romeo, T. (1997). The global regulator CsrA of Escherichia coli is a specific mRNA-binding protein. Journal of Bacteriology, 179(14), 4639–42. Retrieved from http://

www.pubmedcentral.nih.gov/articlerender.fcgi?artid=179305&tool=pmcentrez&rendertype=abstract

Liu, T., Zhang, K., Xu, S., Wang, Z., Fu, H., Tian, B., … Li, W. (2017). Detecting RNA-RNA interactions in E. coli using a modified CLASH method. BMC Genomics, 18(1), 343. https://doi.org/10.1186/s12864-017-3725-3

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. Algorithms for Molecular Biology, 6(1), 26. https://doi.org/10.1186/1748-7188-6-26

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes.Gene, 234, 234(2):187-208. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10395892

Mackie, G. a. (1998). Ribonuclease E is a 5'-end-dependent endonuclease. Nature, 395(October), 720–723. https://doi.org/10.1038/27246

Mackie, G. A. (2012). RNase E: at the interface of bacterial RNA processing and decay. Nature Reviews Microbiology, 11(1), 45–57. https://doi.org/10.1038/nrmicro2930

Maeda, H. (2000). Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Research, 28(18), 3497–3503. https://doi.org/10.1093/nar/28.18.3497

Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. FEBS Letters, 583(24), 3966–3973. https://doi.org/10.1016/j.febslet.2009.10.036

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., & Gottesman, S. (1998). DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. Genetics, 95(October), 12462–12467. https://doi.org/10.1073/pnas.95.21.12462

Majdalani, N., Hernandez, D., & Gottesman, S. (2002). Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. Molecular Microbiology, 46(3), 813–826. https://doi.org/10.1046/j.1365-2958.2002.03203.x

Maki, K., Morita, T., Otaka, H., & Aiba, H. (2010). A minimal base-pairing region of a bacterial small RNA SgrS required for translational repression of ptsG mRNA. Molecular Microbiology, 76(3), 782–792. https://doi.org/10.1111/j.1365-2958.2010.07141.x

Mandin, P., & Gottesman, S. (2010). Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. The EMBO Journal, 29(18), 3094–3107. https://doi.org/10.1038/emboj.2010.179

Mangan, S., & Alon, U. (2003). Structure and function of the feed-forward loop network motif. Proceedings of the National Academy of Sciences of the United States of America, 100(21), 11980–11985. https://doi.org/10.1073/pnas.2133841100

Mascher, T., Helmann, J. D., & Unden, G. (2006). Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases. Microbiology and Molecular Biology Reviews, 70(4), 910–938. https://doi.org/10.1128/MMBR.00020-06

Massé, E., Escorcia, F. E., & Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli, 2374–2383. https://doi.org/10.1101/gad.1127103

Massé, E., Vanderpool, C. K., & Gottesman, S. (2005). Effect of RyhB small RNA on global iron use in Escherichia coli. Journal of Bacteriology, 187(20), 6962–6971. https://doi.org/10.1128/JB.187.20.6962-6971.2005

McCullen, C. A., Benhammou, J. N., Majdalani, N., & Gottesman, S. (2010). Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: Pairing increases translation and protects rpoS mRNA from degradation. Journal of Bacteriology, 192(21), 5559–5571. https://doi.org/10.1128/JB.00464-10

Melamed, S., Peer, A., Faigenbaum-Romm, R., Gatt, Y. E., Reiss, N., Bar, A., … Margalit, H. (2016). Global Mapping of Small RNA-Target Interactions in Resource Global Mapping of Small RNA-Target Interactions in Bacteria. Molecular Cell, 63(5), 884–897. https://doi.org/10.1016/j.molcel.2016.07.026

Michaux, C., Verneuil, N., Hartke, A., & Giard, J.-C. (2014). Physiological roles of small RNA molecules. Microbiology, 160(Pt_6), 1007–1019. https://doi.org/10.1099/mic.0.076208-0

Mika, F., & Hengge, R. (2014). Small RNAs in the control of RpoS, CsgD, and biofilm architecture of Escherichia coli. RNA Biology, 11(5), 494–507. https://doi.org/10.4161/rna.28867

Milo, R. (2002). Network Motifs: Simple Building Blocks of Complex Networks. Science, 298(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

Mira, A., Martín-Cuadrado, A. B., D'Auria, G. & Rodríguez-Valera, F. (2010). The bacterial pan-genome: a new paradigm in microbiology. INTERNATIONAL MICROBIOLOGY 13:45-57. https://doi.org/10.2436/20.1501.01.110

Miyakoshi, M., Chao, Y., & Vogel, J. (2015b). Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. The EMBO Journal, 34(11), 1478–1492. https://doi.org/10.15252/embj.201490546

Miyakoshi, M., Chao, Y., & Vogel, J. (2015a). Regulatory small RNAs from the 3' regions of bacterial mRNAs. Current Opinion in Microbiology, 24, 132–139. https://doi.org/10.1016/j.mib.2015.01.013

Mohanty, B. K., Maples, V. F., & Kushner, S. R. (2004). The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. Molecular Microbiology, 54(4), 905–920. https://doi.org/10.1111/j.1365-2958.2004.04337.x

Molin, C., Jauhiainen, A., Warringer, J., Nerman, O., & Sunnerhagen, P. (2009). mRNA stability changes precede changes in steady-state mRNA amounts during hyperosmotic stress. RNA (New York, N.Y.), 15(4), 600–14. https://doi.org/10.1261/rna.1403509

Moll, I., Afonyushkin, T., Vytvytska, O., & Kaberdin, V. R. (2003). Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs Rna, 9, 1308–1314. https://doi.org/10.1261/rna.5850703.has

Moon, K., & Gottesman, S. (2011). Competition among Hfq-binding small RNAs in Escherichia coli. Molecular Microbiology, 82(6), 1545–1562. https://doi.org/10.1111/j.1365-

2958.2011.07907.x

Morita, T., Maki, K., & Aiba, H. (2005). RNase E-based ribonucleoprotein complexes: Mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. Genes and Development, 19(18), 2176–2186. https://doi.org/10.1101/gad.1330405

Morita, T., Mochizuki, Y., & Aiba, H. (2006). Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. Proceedings of the National Academy of Sciences, 103(13), 4858–4863. https://doi.org/10.1073/pnas.0509638103

Mutalik, V. K., Nonaka, G., Ades, S. E., Rhodius, V. A., & Gross, C. A. (2009). Promoter strength properties of the complete sigma E regulon of Escherichia coli and Salmonella enterica. Journal of Bacteriology, 191(23), 7279–7287. https://doi.org/10.1128/JB.01047-09

Nam, J., Choi, S., & You, B. (2016). Incredible RNA: Dual Functions of Coding and Noncoding. Molecules and Cells, 39(5), 367–374. https://doi.org/10.14348/molcells.2016.0039

Nature, L. T. O. (1990). Letters to nature. Nature, 346(May), 376–379.

Nikado,H. (1994). No Porins and specific difusion channels in bacterial outer membranes. J Biol Chem, 269((6)), 3905–3908.

Neußer, T., Gildehaus, N., Wurm, R., & Wagner, R. (2008). Studies on the expression of 6S RNA from E. coli: Involvement of regulators important for stress and growth adaptation. Biological Chemistry, 389(3), 285–297. https://doi.org/10.1515/BC.2008.023

Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. Microbiology and Molecular Biology Reviews : MMBR, 67(4), 593–656. https://doi.org/10.1128/MMBR.67.4.593

Nitzan, M., Rehani, R., & Margalit, H. (2017). Integration of Bacterial Small RNAs in Regulatory Networks. Annual Review of Biophysics, 46(1), 131–148. https://doi.org/10.1146/annurev-biophys-070816-034058

Opdyke, J. A., Kang, J., & Storz, G. (2004). GadY, a Small-RNA Regulator of Acid Response Genes in Escherichia coli GadY. Journal of Bacteriology, 186(20), 6698–6705. https://doi.org/10.1128/JB.186.20.6698

Otaka, H., Ishikawa, H., Morita, T., & Aiba, H. (2011). PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. Proceedings of the National Academy of Sciences, 108(32), 13059–13064. https://doi.org/10.1073/pnas.1107050108

Overgaard, M., Johansen, J., Møller-Jensen, J., & Valentin-Hansen, P. (2009). Switching off small RNA regulation with trap-mRNA. Molecular Microbiology, 73(5), 790–800. https://doi.org/10.1111/j.1365-2958.2009.06807.x

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. Nature Genetics, 31(1), 69–73. https://doi.org/10.1038/ng869

Paget, M. S. (2015). Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. Biomolecules, 5(3), 1245–1265. https://doi.org/10.3390/biom5031245

Panja, S., Schu, D. J., & Woodson, S. A. (2013). Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. Nucleic Acids Research, 41(15), 7536–7546. https://doi.org/10.1093/nar/gkt521

Papenfort, K., & Vanderpool, C. K. (2015). Target activation by regulatory RNAs in bacteria. FEMS Microbiology Reviews, 39(3), 362–378. https://doi.org/10.1093/femsre/fuv016

Papenfort, K., & Vogel, J. (2014). Small RNA functions in carbon metabolism and virulence of enteric pathogens. Frontiers in Cellular and Infection Microbiology, 4(July), 1–12. https://doi.org/10.3389/fcimb.2014.00091

Papenfort, K., Bouvier, M., Mika, F., Sharma, C. M., & Vogel, J. (2010). Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. Proceedings of the National Academy of Sciences, 107(47), 20435–20440. https://doi.org/10.1073/pnas.1009784107

Papenfort, K., Espinosa, E., Casadesús, J., & Vogel, J. (2015). Small RNA-based feedforward loop with AND-gate logic regulates extrachromosomal DNA transfer in Salmonella. Proceedings of the National Academy of Sciences, 112(34), E4772–E4781. https://doi.org/10.1073/pnas.1507825112

Papenfort, K., Pfeiffer, V., Mika, F., Lucchini, S., Hinton, J. C. D., & Vogel, J. (2006). σE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay. Molecular Microbiology, 62(6), 1674–1688. https://doi.org/10.1111/j.1365-2958.2006.05524.x

Papenfort, K., Said, N., Welsink, T., Lucchini, S., Hinton, J. C. D., & Vogel, J. (2009). Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-aependent small RNA. Molecular Microbiology, 74(1), 139–158. https://doi.org/10.1111/j.1365-2958.2009.06857.x

Peer, A., & Margalit, H. (2011). Accessibility and evolutionary conservation mark bacterial small-RNA target-binding regions. Journal of Bacteriology, 193(7), 1690–1701. https://doi.org/10.1128/JB.01419-10

Peters, J.M., Vangeloff, AD. & Landick R (2011). Bacterial Transcription Terminators: The RNA 3 ′-End Chronicles. J Mol Biol 412(5), 793–813. https://doi.org/10.1016/j.jmb.2011.03.036

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J. C. D., & Vogel, J. (2009). Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nature Structural & Molecular Biology, 16(8), 840–846. https://doi.org/10.1038/nsmb.1631

Picard, F., Dressaire, C., Girbal, L., & Cocaign-Bousquet, M. (2009). Examination of post-transcriptional regulations in prokaryotes by integrative biology. Comptes Rendus - Biologies, 332(11), 958–973. https://doi.org/10.1016/j.crvi.2009.09.005

Plumbridge, J., Bossi, L., Oberto, J., Wade, J. T., & Figueroa-Bossi, N. (2014). Interplay of transcriptional and small RNA-dependent control mechanisms regulates chitosugar uptake in Escherichia coli and Salmonella. Molecular Microbiology, 92(4), 648–658. https://doi.org/10.1111/mmi.12573

Prasse, D., Ehlers, C., Backofen, R., & Schmitz, R. A. (2013). Regulatory RNAs in archaea: first target identification in Methanoarchaea. Biochemical Society Transactions, 41(1), 344–349. https://doi.org/10.1042/BST20120280

Pratt, L. A., Hsing, W., Gibson, K. E., & Silhavy, T. J. (1996). From acids to osmZ: Multiple factors influence synthesis of the OmpF and OmpC porins in Escherichia coli. Molecular Microbiology, 20(5), 911–917. https://doi.org/10.1111/j.1365-2958.1996.tb02532.x

Prévost, K., Desnoyers, G., Jacques, J. F., Lavoie, F., & Massé, E. (2011). Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. Genes and Development, 25(4), 385–396. https://doi.org/10.1101/gad.2001711

Prévost, K., Salvail, H., Desnoyers, G., Jacques, J. F., Phaneuf, É.´, & Massé, E. (2007). The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. Molecular Microbiology, 64(5), 1260–1273. https://doi.org/10.1111/j.1365-2958.2007.05733.x

Radhakrishnan, A., & Green, R. (2016). Connections Underlying Translation and mRNA Stability. Journal of Molecular Biology, 428(18), 3558–3564. https://doi.org/10.1016/j.jmb.2016.05.025

Ragnehed, M. (2009). F Unctional M Agnetic R Esonance, (1121), 1–9. https://doi.org/10.1227/01.NEU.0000158323.91174.3A

Raivio, T. L. (2014). Everything old is new again: An update on current research on the Cpx envelope stress response. Biochimica et Biophysica Acta - Molecular Cell Research, 1843(8), 1529–1541. https://doi.org/10.1016/j.bbamcr.2013.10.018

Rasmussen, A. A., Eriksen, M., Gilany, K., Udesen, C., Franch, T., Petersen, C., & Valentin-Hansen, P. (2005). Regulation of ompA mRNA stability: The role of a small regulatory RNA in growth phase-dependent control. Molecular Microbiology, 58(5), 1421–1429. https://doi.org/10.1111/j.1365-2958.2005.04911.x

Rau, M. H., Bojanovič, K., Nielsen, A. T., & Long, K. S. (2015). Differential expression of small RNAs under chemical stress and fed-batch fermentation in E. coli. BMC Genomics, 16(1), 1051. https://doi.org/10.1186/s12864-015-2231-8

Rau, M. H., Bojanovič, K., Nielsen, A. T., & Long, K. S. (2015). Differential expression of small RNAs under chemical stress and fed-batch fermentation in E. coli. BMC Genomics, 16(1), 1051. https://doi.org/10.1186/s12864-015-2231-8

Reichenbach, B., Maes, A., Kalamorz, F., Hajnsdorf, E., & Görke, B. (2008). The small RNA GlmY acts upstream of the sRNA GlmZ in the activation of glmS expression and is subject to regulation by polyadenylation in Escherichia coli. Nucleic Acids Research, 36(8), 2570–2580. https://doi.org/10.1093/nar/gkn091

Repoila, F., & Darfeuille, F. (2009). Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. Biology of the Cell, 101(2), 117–131. https://doi.org/10.1042/BC20070137

Repoila, F., & Gottesman, S. (2001). Signal Transduction Cascade for Regulation of RpoS : Temperature Regulation of DsrA Signal Transduction Cascade for Regulation of RpoS : Temperature Regulation of DsrA. Growth (Lakeland), 183(13), 4012–4023. https://doi.org/10.1128/JB.183.13.4012

Rhodius, V. A., Suh, W. C., Nonaka, G., West, J., & Gross, C. A. (2006). Conserved and variable functions of the σE stress response in related genomes. PLoS Biology, 4(1), 0043–0059. https://doi.org/10.1371/journal.pbio.0040002

Romby, P., Vandenesch, F., & Wagner, E. G. H. (2006). The role of RNAs in the regulation of virulence-gene expression. Current Opinion in Microbiology, 9(2), 229–236. https://doi.org/10.1016/j.mib.2006.02.005

Romeo, T. (1998). Global regulation by the small RNA-binding protein CsrA and the

non- coding RNA molecule CsrB. Molecular Microbiology, 29(6), 1321–1330. https://doi.org/10.1046/j.1365-2958.1998.01021.x

Romeo, T., Vakulskas C. A. & Babitzke, P (2013). Post-transcriptional regulation on a global scale: function of Csr / Rsm systems, 15(2), 313–324. https://doi.org/10.1111/j.1462-2920.2012.02794.x.

Rowley, G., Spector, M., Kormanec, J., & Roberts, M. (2006). Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. Nature Reviews Microbiology, 4(5), 383–394. https://doi.org/10.1038/nrmicro1394

Saito, S., Kakeshita, H., Nakamura, K., De Lay, N., Schu, D. J., Gottesman, S., … Sachs, G. (2015). Genome-wide detection of novel regulatory RNAs in E . coli. Nucleic Acids Research, 10(1), 1487–1497. https://doi.org/10.1101/gr.119370.110.21

Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P. P. (2011). A ceRNA hypothesis: The rosetta stone of a hidden RNA language? Cell, 146(3), 353–358. https://doi.org/10.1016/j.cell.2011.07.014

Sauer, E. (2013). Structure and RNA-binding properties of the bacterial LSm protein Hfq. RNA Biology, 10(4), 610–8. https://doi.org/10.4161/rna.24201

Schlegel, A., Böhm, A., Lee, S.-J., Peist, R., Decker, K., & Boos, W. (2002). Network regulation of the Escherichia coli maltose system. Journal of Molecular Microbiology and Biotechnology, 4(3), 301–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11931562

Schu, D. J., Zhang, A., Gottesman, S., & Storz, G. (2015). Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. The EMBO Journal, 34(20), 2557–2573. https://doi.org/10.15252/embj.201591569

Schubert, M., Lapouge, K., Duss, O., Oberstrass, F. C., Jelesarov, I., Haas, D., & Allain, F. H.-T. (2007). Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. Nature Structural & Molecular Biology, 14(9), 807–813. https://doi.org/10.1038/nsmb1285

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., … Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature, 473(7347), 337–342. https://doi.org/10.1038/nature10098

Sedlyarova, N., Shamovsky, I., Bharati, B. K., Epshtein, V., Chen, J., Gottesman, S., … Nudler, E. (2016). sRNA-Mediated Control of Transcription Termination in E. coli. Cell, 167(1), 111–121.e13. https://doi.org/10.1016/j.cell.2016.09.004

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. Journal of Bacteriology, 189(23), 8746–8749. https://doi.org/10.1128/JB.01368-07

Shalem, O., Dahan, O., Levo, M., Martinez, M. R., Furman, I., Segal, E., & Pilpel, Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. Molecular Systems Biology, 4(223). https://doi.org/10.1038/msb.2008.59

Shannon, P., Markiel, A., Owen Ozier, 2, Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, (13), 2498–2504. https://doi.org/10.1101/gr.1239303.metabolite

Sharma, C. M., & Vogel, J. (2009). Experimental approaches for the discovery and characterization of regulatory small RNA. Current Opinion in Microbiology, 12(5), 536–546. https://doi.org/10.1016/j.mib.2009.07.006

Sharma, C. M., Darfeuille, F., Plantinga, T. H., & Vogel, J. (2007). Supplementary Material A small RNA regulates multiple ABC transporter mRNAs by targeting C / A-rich elements inside and upstream of ribosome binding sites, 2804–2817. https://doi.org/10.1101/gad.447207

Shihama, B. A. I. (2012). Review Prokaryotic genome regulation : A revolutionary paradigm, 88(9), 485–508.

Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O., & Margalit, H. (2007). Regulation of gene expression by small non-coding RNAs: a quantitative view. Molecular Systems Biology, 3(138), 1–9. https://doi.org/10.1038/msb4100181

Shuman, H. A. (1982). Maltose in, 257(10).

Shine,J. & Dalgarno, L. (1974). The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. Proc Natl Acad Sci U S A. 1974 Apr;71(4):1342-6.

Silhavy, T. J., Ruiz, N., & Kahne, D. (2006). Advances in understanding bacterial outer-membrane biogenesis. Nature Reviews Microbiology, 4(1), 57–66. https://doi.org/10.1038/nrmicro1322

Silhavy, T., Kahne, D., & Walker, S. (2010). The bacterial cell envelope. Cold Spring Harbor Perspectives in Biology, 2(5), 1–16. https://doi.org/10.1101/cshperspect.a000414

Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., … Yusupov, M. (2009). A structural view of translation initiation in bacteria. Cellular and Molecular Life Sciences, 66(3), 423–436. https://doi.org/10.1007/s00018-008-8416-4

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C. M., Rolle, K., Binnewies, T. T., … Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genetics, 4(8). https://doi.org/10.1371/journal.pgen.1000163

Skippington, E., & Ragan, M. A. (2012). Evolutionary dynamics of small RNAs in 27 escherichia coli and shigella genomes. Genome Biology and Evolution, 4(3), 330–345. https://doi.org/10.1093/gbe/evs001

Sklar, J. G., Wu, T., Kahne, D., & Silhavy, T. J. (2007). Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. Genes and Development, 21(19), 2473–2484. https://doi.org/10.1101/gad.1581007

Sledjeski, D., & Gotresman, S. (1995). A small RNA acts as an antisilencer of the H-NS-silenced rcsA gene of Escherichia coli (DsrA/lon/transcription initiation/capsule synthesis). Biochemistry, 92(March 1995), 2003–2007. https://doi.org/10.1073/pnas.92.6.2003

Smirnov, A., Förstner, K. U., Holmqvist, E., Otto, A., Günster, R., Becher, D., … Vogel, J. (2016). Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. Proceedings of the National Academy of Sciences, 113(41), 11591–11596. https://doi.org/10.1073/pnas.1609981113

Smirnov, A., Wang, C., Drewry, L. L., & Vogel, J. (2017). Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. The EMBO Journal, 36(8), 1029–

1045. https://doi.org/10.15252/embj.201696127

Sneppen, K., Pedersen, S., Krishna, S., Dodd, I., & Semsey, S. (2010). Economy of operon formation: Cotranscription minimizes shortfall in protein complexes. mBio, 1(4), 3–5. https://doi.org/10.1128/mBio.00177-10

Soper, T., Mandin, P., Majdalani, N., Gottesman, S., & Woodson, S. a. (2010). Positive regulation by small RNAs and the role of Hfq. Pnas, 107(21), 2–7. https://doi.org/10.1073/pnas.1004435107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1004435107

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., … Von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Research, 43(D1), D447–D452. https://doi.org/10.1093/nar/gku1003

Szmelcman, S., & Hofnung, M. (1975). Maltose transport in Escherichia coli K-12: involvement of the bacteriophage lambda receptor. Journal of Bacteriology, 124(1), 112–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=235871&tool=pmcentrez&rendertype=abstract

Szmelcman, S., Schwartz, M., Silhavy, T. J., & Boos, W. (1976). Maltose transport in Escherichia coli K12. European Journal of Biochemistry, 19(1), 13–19. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1976.tb10383.x/full

Thomason, M. K., & Storz, G. (2010). Bacterial antisense RNAs: How many are there and what are they doing?. Annu Rev Genet. 44, 167-188. https://doi.org/10.1146/annurev-genet-102209-163523

Thomason, M. K., Bischler, T., Eisenbart, S. K., Förstner, K. U., Zhang, A., Herbig, A., … Storza, G. (2015). Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. Journal of Bacteriology, 197(1), 18–28. https://doi.org/10.1128/JB.02096-14

Thompson, K. M., Rhodius, V. A., & Gottesman, S. (2007). σE regulates and is regulated by a small RNA in Escherichia coli. Journal of Bacteriology, 189(11), 4243–4256. https://doi.org/10.1128/JB.00020-07

Tramonti, A., De Canio, M., & De Biase, D. (2008). GadX/GadW-dependent regulation of the Escherichia coli acid fitness island: Transcriptional control at the gadY-gadW divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. Molecular Microbiology, 70(4), 965–982. https://doi.org/10.1111/j.1365-2958.2008.06458.x

Travers, A., & Muskhelishvili, G. (2005). DNA supercoiling — a global transcriptional regulator for enterobacterial growth? Nature Reviews Microbiology, 3(2), 157–169. https://doi.org/10.1038/nrmicro1088

Travis, A. J., Moody, J., Helwak, A., Tollervey, D., & Kudla, G. (2013). Ligation and Sequencing of Hybrids) Data Q. Methods, 65, 263–273. https://doi.org/10.1016/j.ymeth.2013.10.015

Tree, J. J., Granneman, S., McAteer, S. P., Tollervey, D., & Gally, D. L. (2014). Identification of Bacteriophage-Encoded Anti-sRNAs in Pathogenic Escherichia coli. Molecular Cell, 55(2), 199–213. https://doi.org/10.1016/j.molcel.2014.05.006

Trotochaud, A. E., & Wassarman, K. M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. Nature Structural & Molecular Biology, 12(4), 313–9.

https://doi.org/10.1038/nsmb917

Udekwu, K. I., Darfeuille, F., Vogel,J., Reimegard, J., Holmqvist, E. & Wagner, E. G. H. (2005). Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. Genes and Development, 19(9), 2355–2366. https://doi.org/10.1101/gad.354405

Udekwu, K. I., & Wagner, E. G. H. (2007). Sigma E controls biogenesis of the antisense RNA MicA. Nucleic Acids Research, 35(4), 1279–1288. https://doi.org/10.1093/nar/gkl1154

Ulrich, L. E., Koonin, E. V, & Zhulin, I. B. (2009). NIH Public Access, 13(2), 52–56. https://doi.org/10.1016/j.tim.2004.12.006.One-component

Urban, J. H., & Vogel, J. (2007). Translational control and target recognition by Escherichia coli small RNAs in vivo. Nucleic Acids Research, 35(3), 1018–1037. https://doi.org/10.1093/nar/gkl1040

Urban, J. H., & Vogel, J. (2009). Riboswitches, 540(1), 301–319. https://doi.org/10.1007/978-1-59745-558-9

Ureta, A., Endres, R. G., Wingreen, N. S. & Silhavy, T. J. (2007). Kinetic Analysis of the Assembly of the Outer Membrane Protein LamB in Escherichia coli Mutants Each Lacking a Secretio or Targeting Factor in a Different Cellular Compartment. J. Bacteriol 189 (2), 446-454. https://doi.org/10.1128/JB.01103-06

Van Assche, E., Van Puyvelde, S., Vanderleyden, J., & Steenackers, H. P. (2015). RNA-binding proteins involved in post-transcriptional regulation in bacteria. Frontiers in Microbiology, 6(MAR), 1–16. https://doi.org/10.3389/fmicb.2015.00141

Van Nues, R. W., Castro-Roa, D., Yuzenkova, Y., & Zenkin, N. (2015). Ribonucleoprotein particles of bacterial small non-coding RNA IsrA (IS61 or McaS) and its interaction with RNA polymerase core may link transcription to mRNA fate. Nucleic Acids Research, 44(6), 2577–2592. https://doi.org/10.1093/nar/gkv1302

van Nues, R., Schweikert, G., de Leau, E., Selega, A., Langford, A., Franklin, R., … Granneman, S. (2017). Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress. Nature Communications, 8(1), 12. https://doi.org/10.1038/s41467-017-00025-5

Vanderpool, C. K. (2007). Physiological consequences of small RNA-mediated regulation of glucose-phosphate stress. Current Opinion in Microbiology, 10(2), 146–151. https://doi.org/10.1016/j.mib.2007.03.011

Večerek, B., Moll, I., & Bläsi, U. (2007). Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. The EMBO Journal, 26(4), 965–975. https://doi.org/10.1038/sj.emboj.7601553

Vogel, J. (2009). A rough guide to the non-coding RNA world of Salmonella. Molecular Microbiology, 71(1), 1–11. https://doi.org/10.1111/j.1365-2958.2008.06505.x

Vogel, J., & Luisi, B. F. (2011). Hfq and its constellation of RNA. Nature Reviews Microbiology, 9(8), 578–589. https://doi.org/10.1038/nrmicro2615

Vogel, J., Bartels, V., Tang, T. H., Churakov, G., Slagter-Jäger, J. G., Hüttenhofer, A., & Wagner, E. G. H. (2003). RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Research, 31(22), 6435–6443. https://doi.org/10.1093/nar/gkg867

Vogt, S. L., Evans, A. D., Guest, R. L., & Raivio, T. L. (2014). The Cpx envelope stress response regulates and is regulated by small noncoding RNAs. Journal of Bacteriology, 196(24), 4229–4238. https://doi.org/10.1128/JB.02138-14

Wadler, C. S., & Vanderpool, C. K. (2007). A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proceedings of the National Academy of Sciences, 104(51), 20454–20459. https://doi.org/10.1073/pnas.0708102104

Wagner, E. G. H. (2013). Cycling of RNAs on Hfq. RNA Biology, 10(4), 619–626. https://doi.org/10.4161/rna.24044

Wagner, L. A., Gesteland, R. F., Dayhuff, T. J., & Weiss, R. B. (1994). An efficient shine-dalgarno sequence but not translation is necessary for lacZ mRNA stability in Escherichia coli. Journal of Bacteriology, 176(6), 1683–1688.

Walker, M., Kublin, J. G., & Zunt, J. R. (2009). NIH Public Access, 42(1), 115–125. https://doi.org/10.1086/498510.Parasitic

Washburn, R. S., & Gottesman, M. E. (2015). Regulation of transcription elongation and termination. Biomolecules, 5(2), 1063–1078. https://doi.org/10.3390/biom5021063

Wassarman, K. M. (2012). 6S RNA: A regulator of transcription. Regulatory RNAs in Prokaryotes, 109–130. https://doi.org/10.1007/978-3-7091-0218-3_6

Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G., & Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. Genes & Development, 15, 1637–1651. https://doi.org/10.1101/gad.901001.identified

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics, 25(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Waters, L. S., & Storz, G. (2009). Regulatory RNAs in Bacteria. Cell, 136(4), 615–628. https://doi.org/10.1016/j.cell.2009.01.043

Waters, S. A., McAteer, S. P., Kudla, G., Pang, I., Deshpande, N. P., Amos, T. G., … Tree, J. J. (2017). Small RNA interactome of pathogenic E. coli revealed through crosslinking of RNase E. The EMBO Journal, 36(3), 374–387. https://doi.org/10.15252/embj.201694639

Webb, S., Hector, R. D., Kudla, G., & Granneman, S. (2014). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. Genome Biology, 15(1), R8. https://doi.org/10.1186/gb-2014-15-1-r8

Weel-Sneve, R., Kristiansen, K. I., Odsbu, I., Dalhus, B., Booth, J., Rognes, T., … Bjørås, M. (2013). Single Transmembrane Peptide DinQ Modulates Membrane-Dependent Activities. PLoS Genetics, 9(2). https://doi.org/10.1371/journal.pgen.1003260

Weilbacher, T., Suzuki, K., Dubey, A. K., Wang, X., Gudapaty, S., Morozov, I., … Romeo, T. (2003). A novel sRNA component of the carbon storage regulatory system of Escherichia coli. Molecular Microbiology, 48(3), 657–670. https://doi.org/10.1046/j.1365-2958.2003.03459.x

Wilson, K. S., & Hippel, P. H. Von. (1995). Transcription termination at intrinsic terminators: The role of the RNA hairpin (Escherichia coli/RNA polymerase/rho-independent termination). Biochemistry, 92(September), 8793–8797. https://doi.org/10.1073/pnas.92.19.8793

Wright, P. R., Georg, J., Mann, M., Sorescu, D. A., Richter, A. S., Lott, S., … Backofen, R. (2014). CopraRNA and IntaRNA: Predicting small RNA targets, networks and interaction domains. Nucleic Acids Research, 42(W1), 119–123. https://doi.org/10.1093/nar/gku359

Wright, P. R., Richter, A. S., Papenfort, K., Mann, M., Vogel, J., Hess, W. R., … Georg, J. (2013). Comparative genomics boosts target prediction for bacterial small RNAs. Proceedings of the National Academy of Sciences, 110(37), E3487–E3496. https://doi.org/10.1073/pnas.1303248110

Wu, X., & Bartel, D. P. (2017). KpLogo: Positional k -mer analysis reveals hidden specificity in biological sequences. Nucleic Acids Research, 45(W1), W534–W538. https://doi.org/10.1093/nar/gkx323

Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G., & Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. Molecular Microbiology, 50(4), 1111–1124. https://doi.org/10.1046/j.1365-2958.2003.03734.x