

# Identification and functional characterization of bacterial small RNAs associating with the RNA chaperone Hfq



## Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

vorgelegt von

Michaela Huber

München, 2020



Diese Dissertation wurde angefertigt  
im Zeitraum von September 2016 bis September 2020  
unter der Leitung von Herrn Prof. Dr. Kai Papenfort  
an der Fakultät für Biologie, Department I,  
an der Ludwig-Maximilians-Universität München.

Erstgutachter: Prof. Dr. Kai Papenfort

Zweitgutachter: Prof. Dr. Simon Ringgaard

Tag der Abgabe: 05.10.2020

Tag der mündlichen Prüfung: 21.01.2021



# Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt wurde. Des Weiteren erkläre ich hiermit, dass die Dissertation weder ganz noch in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist, und dass ich zu keinem früheren Zeitpunkt versucht habe, eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

I declare that I have authored this thesis independently and that I have not used other than the declared sources. As well I declare that the dissertation has not been presented to another examination board, neither entirely nor in parts, and that I have not submitted or defended a dissertation previously without success.

München, 30.09.2020

Michaela Huber



# Table of Contents

<b>Publications and Contributions</b> .....	<b>IX</b>
<b>Summary</b> .....	<b>XV</b>
<b>Zusammenfassung</b> .....	<b>XVII</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 RNA-based regulation of gene expression in bacteria .....	1
1.2 <i>Trans</i> -acting small regulatory RNAs .....	3
1.2.1 Functional characterization of <i>trans</i> -acting sRNAs .....	3
1.2.2 Physiological roles of sRNAs and their impact on regulatory networks .....	6
1.3 Proteins required for regulation by sRNAs.....	10
1.3.1 The RNA chaperone Hfq.....	10
1.3.2 Other RNA-binding proteins.....	12
1.3.3 The ribonuclease RNase E .....	13
1.4 Characterization and identification of Hfq-binding sRNAs.....	15
1.5 Model organisms used in this study.....	17
1.5.1 <i>Vibrio cholerae</i> .....	18
1.5.2 <i>Photobacterium laumondii</i> and <i>Xenorhabdus szentirmaii</i> .....	19
1.6 Aim and scope of the study .....	20
<b>2 Switching fatty acid metabolism by an RNA-controlled feed forward loop</b> (publication) .....	<b>23</b>
<b>3 Gene autoregulation by 3' UTR-derived bacterial small RNAs</b> (publication).....	<b>59</b>
<b>4 Symbiosis, virulence and natural products biosynthesis in entomopathogenic</b> <b>bacteria are regulated by a small RNA</b> (accepted manuscript).....	<b>115</b>
<b>5 RIL-seq reveals a quorum sensing small RNA sponge</b> (manuscript in preparation)	<b>155</b>
<b>6 Discussion</b> .....	<b>185</b>
6.1 Detecting sRNAs and their targets with high-throughput approaches .....	185
6.1.1 Global detection of Hfq-binding sRNAs via co-immunoprecipitation .....	185
6.1.2 Identification and validation of target RNAs .....	187
6.1.3 Integration of multiple data sets .....	190
6.2 Target regulation by Hfq-binding sRNAs .....	193
6.3 Origin and evolution of sRNAs and the role of Hfq and RNase E .....	196
6.4 ArcZ is a highly conserved sRNA .....	198

## TABLE OF CONTENTS

6.5 Integration of sRNAs in regulatory circuits.....	200
6.5.1 Autoregulatory feedback regulation at the RNA level.....	201
6.5.2 A feed forward loop allowing nuanced regulation of <i>V. cholerae</i> 's fatty acid metabolism .....	202
6.6 Summary and perspective.....	205
<b>References .....</b>	<b>207</b>
<b>Abbreviations .....</b>	<b>220</b>
<b>Acknowledgements .....</b>	<b>222</b>
<b>Curriculum vitae .....</b>	<b>223</b>

# Publications and Contributions

Publications and manuscripts originating from this thesis:

## Chapter 2:

**Michaela Huber**, Kathrin S. Fröhlich, Jessica Radmer, and Kai Papenfort (2020). “Switching fatty acid metabolism by an RNA-controlled feed forward loop.” *Proc Natl Acad Sci U S A* 117(14): 8044-8054.

## Chapter 3:

Mona Hoyos, **Michaela Huber**, Konrad U. Förstner, and Kai Papenfort (2020). “Gene autoregulation by 3' UTR-derived bacterial small RNAs.” *eLife* 9, e58836.

## Chapter 4:

Nick Neubacher\*, Nicholas J. Tobias\*, **Michaela Huber\***, Xiaofeng Cai, Timo Glatzer, Sacha J. Pidot, Timothy P. Stinear, Anna Lena Lütticke, Kai Papenfort, and Helge B. Bode (2020). “Symbiosis, virulence and natural products biosynthesis in entomopathogenic bacteria are regulated by a small RNA.”

Manuscript accepted for publication in *Nature Microbiology*.

\* These authors contributed equally.

## Chapter 5:

**Michaela Huber**, Mona Hoyos, Sahar Melamed, Alexandre Vicente, Gisela Storz, and Kai Papenfort. “RIL-seq reveals a quorum sensing small RNA sponge.”

Manuscript in preparation.

## Manuscript in preparation not presented in this thesis:

Laura Vogt, Nikolai Peschek, Anna Schäpers, **Michaela Huber**, Kai Papenfort, and Kathrin S. Fröhlich. “Co-immunoprecipitation with Hfq reveals a novel iron-responsive small RNA in *Caulobacter crescentus*.”

## Contributions to publications and manuscripts presented in this thesis:

### Chapter 2:

MH and KP conceptualized the study. MH performed the majority of experiments and data analyses (Fig. 1A-E, Fig. 2A-E, Fig. 3A, Fig. 4A-E, Fig. 5C, Fig. 6B-E, Fig. S1A-B, Fig. S2A-F, Fig. S3A-C, Fig. S4A, Fig. S5A-E, Fig. S6A-H). JR contributed to plasmid and strain construction, to Northern blot probing of Hfq-binding sRNAs and to the investigation of the transcriptional regulation of FarS and the identification of the RNase E cleavage site in *fabB-farS* (Fig. 1E, Fig. 2E, Fig. 3B, Fig. S1B, Fig. S3B). MH and JR performed experiments to analyze FarS target regulation over growth (Fig. 5A-B). KSF performed radioactive labelling of in vitro RNA transcripts and FarS structure probing (Fig. S2A-F, Fig. S4B). MH was assisted by the research student Kawtar Tiamani. MH constructed the figures, KP and MH wrote the manuscript.

.....  
Michaela Huber

.....  
Prof. Dr. Kai Papenfort

**Chapter 3:**

MHo and KP conceptualized the study. MHo performed the majority of experiments and data analyses (Fig. 1A, Fig. 1-S1, Fig. 1-S4, Fig. 1-S5, Fig. 2A-C, Fig. 2-S1, Fig. 3B, D, Fig. 3-S1C-D, Fig. 3-S2C, Fig. 3-S3, Fig. 4, Fig. 4-S1, Fig. 5, Fig. 6, Fig. 7, Fig. 7-S1A-B, Fig. 7-S2, Fig. 8, Fig. 8-S1, Fig. 9). MHu contributed by plasmid and strain construction, by analyzing Hfq dependence of OppZ and CarZ through Hfq co-immunoprecipitation and sRNA stability experiments, by Northern blot probing, by measuring OppZ repression strength, and by the initial characterization of CarZ (Fig. 1-S6, Fig. 2D, Fig. 2-S2, Fig. 3C, Fig. 5D, Fig. 7-S1C). MHo and MHu performed OppZ pulse expression and analyzed sequencing data (Fig. 3A, Fig. 3-S1A-B). KUF analyzed TIER-seq data (Fig. 1B-D, Fig. 1-S2, Fig. 1-S3). MHo constructed the figures, KP and MHo wrote the original draft of the manuscript, MHu helped finalizing the manuscript.

.....  
Michaela Huber

.....  
Prof. Dr. Kai Papenfort

**Chapter 4:**

NN, NT, MH, KP and HB conceptualized the study. NN, NT and MH performed the majority of experiments and data analyses. XC generated the *hfq::3XFLAG* strain, MH and XC performed RIP-seq experiments, NT analyzed RIP-seq data (Fig. 1b-c, Suppl. Fig. 3, Suppl. Tables 4 and 5). NN and NT performed and analyzed all other RNA-sequencing experiments (Fig. 4a-b, Suppl. Fig. 2a-b, Suppl. Tables 1, 2, 7, 8 and 12). MH performed Northern blot probing of sRNA candidates (Ext. Data Figs. 1 and 2). NN performed transposon mutagenesis and analysis (Ext. Data Fig. 3, Suppl. Fig. 1), analyses of SM production (Fig. 2b-h, Fig. 3e, Fig. 4c, Suppl. Fig. 4a-b) and nematode development assay (Ext. Data Fig. 4), and constructed all mutant strains, unless stated otherwise. MH performed Northern blot probing to analyze expression of *ArcZ* in *P. laumondii* and in *X. szentirmaii* (Fig. 2a, Fig. 3d). NT analyzed conservation of *ArcZ* (Fig. 1a) and predicted *ArcZ-hexA* base-pairing interaction (Suppl. Table 10). MH analyzed direct regulation of *hexA* by *ArcZ*, including plasmid and strain construction (Fig. 3a-c, Suppl. Fig. 5). AL generated the *X. szentirmaii*  $\Delta$ *arcZ* strain and helped with the nematode development assay (Ext. Data Fig. 4). TG performed proteome analysis (Suppl. Table 11). SP and TS sequenced and analyzed transposon insertion mutants (Ext. Data Fig. 3). NN, NT and MH constructed the figures and wrote the manuscript, KP and HB commented on the manuscript.

.....  
 Michaela Huber

.....  
 Prof. Dr. Kai Papenfort

.....  
 Nick Neubacher

.....  
 Nicholas Tobias

**Chapter 5:**

MHu, MHo and KP conceptualized the study. MHu and SH performed RIL-seq experiments and analyzed the data (Fig. 1 A and B). SH was assisted by the bioinformatician Hongen Zhang. GS supervised the RIL-seq study. MHu performed RIL-seq validation (Fig. 1C) and was assisted by the research students Jessica Radmer and Kawtar Tiamani. MHu analyzed expression, conservation and base-pairing interactions of QrrS (Fig. 2). MHu and MHo performed transcript stability experiments to examine sponge function of QrrS (Fig. 3). MHo and AV identified the transcription factor VCA0830 (Fig. 4B), MHu and MHo further investigated transcriptional regulation of QrrS using a transcriptional reporter fusion (Fig. 4C) and analyzed expression of VCA0830 (Fig. 4D). MHu performed the bioluminescence assay (Fig. 5). MHu constructed the figures and wrote the manuscript.

.....  
Michaela Huber

.....  
Prof. Dr. Kai Papenfort



# Summary

Bacteria constantly fine-tune their gene expression in response to environmental changes and in this process gene regulation at the post-transcriptional level plays a major role. In the last two decades, small regulatory RNAs (sRNAs), which act by intermolecular base-pairing with target RNAs, have emerged as key players here. In Gram-negative bacteria, these sRNAs often depend on the global RNA chaperone Hfq, stabilizing sRNAs and mediating their interaction with target mRNAs. It has been shown in various microbial species that Hfq-binding sRNAs regulate hundreds of transcripts in the cell and play pivotal roles in diverse bacterial processes, such as regulation of metabolic pathways, quorum sensing and virulence.

In this thesis, RNA-ligands of Hfq were identified in the human pathogen *Vibrio cholerae* by performing RIP-seq (RNA immunoprecipitation followed by RNA-sequencing) analysis. With this approach, dozens of sRNAs and several hundreds of mRNAs were detected to associate with Hfq. The publication presented in chapter 2 focuses on one representative example, which was one of the most abundant Hfq-binding sRNAs discovered in these experiments. The sRNA, named FarS (fatty acid regulated sRNA), post-transcriptionally represses two paralogous genes of the fatty acid degradation pathway. It is part of a novel type of RNA-based feed forward loop, regulating the transition between fatty acid biosynthesis and degradation in *V. cholerae*.

In the publication presented in chapter 3, another genome-wide approach, TIER-seq (transiently inactivating an endoribonuclease followed by RNA-seq), was used to identify cleavage sites of RNase E, a major ribonuclease in Gram-negative bacteria. With the resulting data set, *V. cholerae* sRNAs were detected that are processed from the 3' UTR of mRNAs. Detailed studies on two of these sRNAs, OppZ and CarZ, revealed that these act as autoregulatory elements and represent a hitherto unknown concept of feedback autoregulation at the RNA level.

In addition to investigating sRNAs in the model organism *V. cholerae*, one chapter of this thesis (chapter 4) is dedicated to the discovery of Hfq-binding sRNAs in the entomopathogenic bacterium *Photobacterium laumondii*, an organism where only little is known about post-transcriptional gene regulation by sRNAs. Our results here further show that the Hfq-binding sRNA ArcZ, previously described in *Escherichia coli* and *Salmonella enterica*, represses a master regulator of secondary metabolite production in *P. laumondii*. This uncovers a so far unknown function of ArcZ and provides novel insights into sRNA-based gene regulation of secondary metabolites in this species.

## SUMMARY

In chapter 5, a recently developed method was used to reveal RNA-RNA interactions in *V. cholerae* at a global scale. With this technique, called RIL-seq (RNA interaction by ligation and sequencing), thousands of interactions were predicted, thereby facilitating enormously the identification of sRNA targets and investigation of their regulatory roles. Here, a novel sRNA was found that base-pairs with the well-studied quorum sensing sRNAs and acts as a central player in the quorum sensing circuit of *V. cholerae*.

In summary, the thesis provides an important contribution to the study of Hfq-binding sRNAs with focus on the model organism *V. cholerae*, but also gives valuable insights into sRNA-based regulation in entomopathogenic bacteria. Furthermore, the impact of the findings is not restricted to the organisms under study, but includes general principles of the functioning of Hfq-binding sRNAs in bacteria.

# Zusammenfassung

Bakterien stimmen fortwährend ihre Genexpression genau auf Veränderungen in ihrer Umwelt ab und in diesem Prozess spielt Genregulation auf post-transkriptioneller Ebene eine bedeutsame Rolle. In den letzten zwei Jahrzehnten haben sich hier kleine regulatorische RNAs (sRNAs für engl. **s**mall **R**NAs), die über intermolekulare Basenpaarung mit Ziel-RNAs agieren, als wesentliche Faktoren herausgestellt. In Gram-negativen Bakterien sind diese sRNAs oft von dem globalen RNA-Chaperone Hfq abhängig, welches sRNAs stabilisiert und ihre Interaktion mit Ziel-RNAs vermittelt. In verschiedenen Mikroben-Arten wurde gezeigt, dass Hfq-bindende sRNAs hunderte von Transkripten in der Zelle regulieren und bei diversen bakteriellen Prozessen, wie zum Beispiel der Regulation von Stoffwechselwegen, Quorum Sensing und Virulenz, Schlüsselrollen spielen.

In der vorliegenden Arbeit wurden RNA-Bindepartner von Hfq in *Vibrio cholerae*, einem Humanpathogen, mittels RIP-seq (RNA-Immunopräzipitation gefolgt von RNA-Sequenzierung)-Analyse identifiziert. Mit diesem Ansatz wurden dutzende sRNAs und mehrere hunderte mRNAs entdeckt, die eine Verbindung mit Hfq eingehen. Die Veröffentlichung in Kapitel 2 konzentriert sich auf ein repräsentatives Beispiel, welches eine der abundantesten Hfq-bindenden sRNAs in diesen Experimenten war. Diese sRNA, als FarS (für engl. **f**atty **a**cid **r**egulated **s**RNA) bezeichnet, reprimiert post-transkriptionell zwei paraloge Gene des Fettsäureabbaus. Sie ist Teil eines neuen Typus von RNA-basierter Feed Forward-Steuerung, welche den Wechsel zwischen Fettsäurebiosynthese und -abbau in *V. cholerae* reguliert.

In der Publikation in Kapitel 3 wurde ein weiterer Genom-weiter Ansatz, TIER-seq (vorrübergehende Inaktivierung einer Endoribonuklease gefolgt von RNA-Sequenzierung), genutzt, um Schnittstellen der RNase E, einer wichtigen Ribonuklease in Gram-negativen Bakterien, zu identifizieren. Mit dem resultierenden Datensatz wurden *V. cholerae* sRNAs aufgedeckt, die aus dem 3' Ende von mRNAs herausprozessiert werden. Detaillierte Untersuchungen zu zwei dieser sRNAs, OppZ and CarZ, zeigen, dass diese als selbstregulierende Elemente fungieren und ein bisher unbekanntes Konzept von Feedback-Selbstregulation auf RNA-Ebene darstellen.

Zusätzlich zu Untersuchungen zu sRNAs im Model-Organismus *V. cholerae* ist ein Kapitel der Arbeit (Kapitel 4) der Entdeckung von Hfq-bindenden sRNAs im insektenpathogenen Bakterium *Photobacterium laumondii* gewidmet, einem Organismus, in dem nur wenig über post-transkriptionelle Regulation durch sRNAs bekannt ist. Unsere Ergebnisse hier zeigen darüber hinaus, dass die Hfq-bindende sRNA ArcZ, die zuvor in *Escherichia coli* und *Salmonella enterica* beschrieben wurde, einen Hauptregulator für die

## ZUSAMMENFASSUNG

Produktion von sekundären Stoffwechselprodukten in *P. laumondii* reprimiert. Dies deckt eine bisher unbekannte Funktion von ArcZ auf und bietet neue Einblicke in sRNA-basierte Genregulation von sekundären Stoffwechselprodukten in dieser Spezies.

In Kapitel 5 wurde eine vor kurzem entwickelte Methode verwendet, um RNA-RNA-Interaktionen in *V. cholerae* im globalen Maßstab aufzudecken. Mit dieser Technik, die RIL-seq (RNA-Interaktion durch Ligation und Sequenzierung) genannt wird, wurden tausende von Interaktionen vorhergesagt, was die Identifizierung von Zielen der sRNAs und die Untersuchung ihrer regulatorischen Rollen enorm vereinfacht. Hier wurde eine neue sRNA gefunden, die mit den gut untersuchten Quorum Sensing - sRNAs eine Basenpaarung eingeht, und als zentraler Faktor im Quorum Sensing - System von *V. cholerae* agiert.

Insgesamt leistet die Arbeit einen wichtigen Beitrag für die Forschung an Hfq-bindenden sRNAs mit Schwerpunkt auf dem Modelorganismus *V. cholerae*, gibt aber auch wertvolle Einblicke in sRNA-basierte Regulation in insektenpathogenen Bakterien. Darüber hinaus ist die Bedeutung der Erkenntnisse nicht auf die verwendeten Organismen beschränkt, sondern umfasst allgemeine Prinzipien der Funktionsweise Hfq-bindender sRNAs in Bakterien.

# 1 Introduction

## 1.1 RNA-based regulation of gene expression in bacteria

For a long time, textbooks presented bacterial gene regulation as a primarily protein-based mechanism with transcription factors playing the key role to activate and repress genes. The function of RNA was considered to be either messenger RNA, transfer RNA or ribosomal RNA, all three interacting to convert information from the universal DNA code into functional proteins. With the discovery of RNA molecules acting as regulators, the understanding of gene expression and regulation has changed profoundly. It is now clear that RNA fulfills regulatory functions to a similar scope as transcription factors do (1). The first bacterial regulatory RNAs have been identified serendipitously in the 1980s (2). Since then, research on the discovery and characterization of bacterial regulatory RNAs has expanded tremendously. The growing number of fully sequenced bacterial genomes has facilitated this research enormously. Moreover, advances in and increasing availability of global techniques based on high-throughput RNA sequencing especially in the last ten years have contributed to rapid developments and discoveries in this emerging field (3). Regulatory RNAs can be found in all three domains of life, and research on eukaryotic riboregulators has started to flourish in parallel to a similar extent (4).

The domain of bacteria consists of highly diverse organisms inhabiting an extremely broad range of ecological niches, including complex symbiotic and parasitic relationships with other organisms. Changing environmental conditions require the ability of fast adaptation based on a rapid and tight regulation of gene expression. It is now well established that regulatory RNAs are involved in various aspects of bacterial physiology and allow dynamic responses to environmental cues (5). Bacterial regulatory RNAs have turned out to be versatile tools of gene regulation acting by different mechanisms. Various types of riboregulators, which have been subject to intense study, will be presented in brief in the following.

The so called RNA thermometers (RNATs) are elements in the 5' UTR of mRNAs which alter their structure in response to temperature. This allows a rapid change in gene expression when bacteria experience a temperature switch. Usually, a stable hairpin structure masks the ribosome-binding site (RBS) at lower temperatures and prevents translation. A shift to higher temperatures leads to a thermodynamically induced unfolding of the inhibitory structure, and in consequence, the RBS becomes accessible allowing the mRNA to be translated (6, 7). For many pathogenic bacteria, this type of regulation can be

## 1 INTRODUCTION

especially relevant when colonizing the host. The increase in temperature upon entering the host can serve as an important trigger to induce virulence-related genes (7, 8). An example is the *toxT* mRNA of the human pathogen *Vibrio cholerae*. ToxT is a transcriptional regulator important for the bacterium's virulence program (9, 10). In aquatic environments, *V. cholerae*'s natural habitat, translation is usually blocked by an inhibitory hairpin structure in the *toxT* 5' UTR sequestering the Shine-Dalgarno (SD) sequence. Upon uptake into the human host, a temperature increase to 37°C results in a conformational change of the inhibitory structure, and subsequently the ribosome is able to access the mRNA (11).

Structural changes are not only the basis for RNATs but also for riboswitches, another type of RNA regulators usually located in the 5' UTR of mRNAs. These short and highly structured elements are characterized by regulating gene expression in response to binding of a small molecule ligand. Riboswitches consist of an aptamer domain and an expression platform. Binding of a specific small molecule to the aptamer domain results in a conformational change in the expression platform, which influences and alters expression of the downstream coding sequence (12). Riboswitches are present in a large diversity of bacterial lineages and can be categorized by their structural features and the ligand they are binding (13).

The most extensively studied type of bacterial RNA regulators are small base-pairing RNAs. These small RNAs (sRNAs) can be divided into two groups: *cis*- and *trans*-acting. *Cis*-acting, also referred to as *cis*-encoded, sRNAs are located on the opposite DNA strand of the mRNA they are regulating, and are typically characterized by an extensive region of perfect complementarity with their target (14). Interestingly, many *cis*-encoded sRNAs are located on mobile genetic elements, such as plasmids, transposons and phages, regulating mainly replication control, transposition and the switch between lysis and lysogeny, but can be found on the chromosome as well (15). For instance, a well-studied group of *cis*-encoded sRNAs are those acting as antitoxins in type I toxin-antitoxin systems. Here, the antitoxin sRNA base-pairs to an mRNA encoding for a toxin and regulates its expression (16). In contrast to *cis*-acting sRNAs, *trans*-acting or *trans*-encoded sRNAs regulate targets whose genes are located at a distant genomic position (17). These riboregulators will be discussed in detail in section 1.2.

Conceptually different to base-pairing sRNAs are regulatory sRNAs that act by modulating protein activity. The prototypic example is the Csr (carbon storage regulator) sRNA family binding to the CsrA protein. CsrA is a global RNA-binding protein influencing translation and transcription attenuation of a broad range of mRNAs in various bacterial species (reviewed in (18)). By binding multiple CsrA molecules, the Csr sRNAs act as sponges and sequester the protein from its mRNA targets (19, 20).

Research in the last decades has shown that RNA-based regulation in bacteria is much more common and diverse than previously assumed (3). Although the crucial role of transcription factors for gene regulation is irrefutable, it is now well understood that regulatory RNA elements are equally important factors to guarantee the functioning of complex gene regulation. The production of short and untranslated regulatory RNA elements seems rather cost-effective for the cell and the regulatory outcome can be reached very fast (17), as expression of mRNAs can be switched on and off rapidly by the diverse mechanisms described above.

## 1.2 *Trans*-acting small regulatory RNAs

### 1.2.1 Functional characterization of *trans*-acting sRNAs

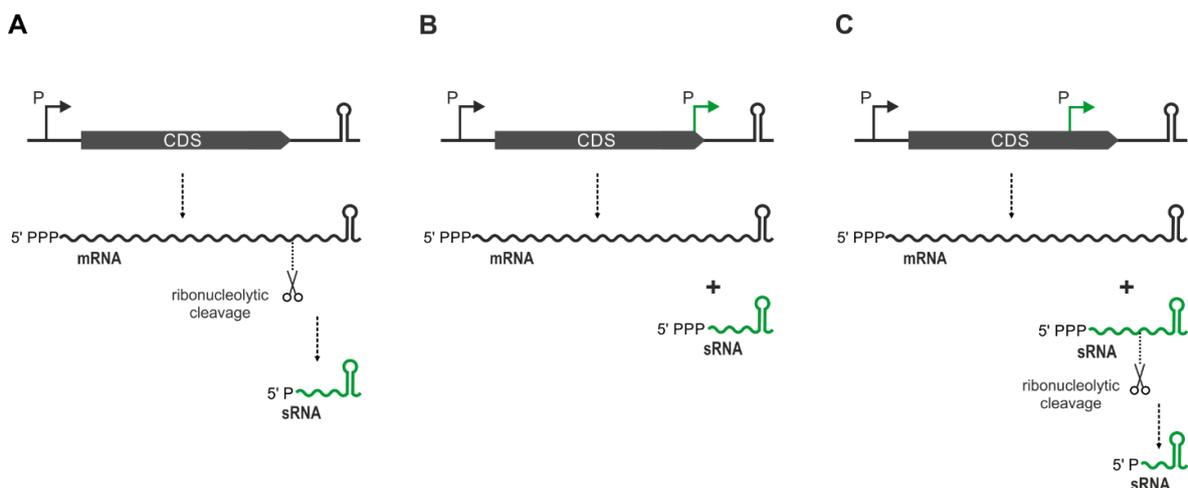
Bacterial small regulatory RNAs that act in *trans* are characterized by a limited and imperfect base-pairing to target RNAs encoded far away from the sRNA genes (17). The base-pairing region typically encompasses a stretch of 10 to 25 nucleotides (17), whereas it has been shown that as few as 4-6 core base-pairs can be sufficient for regulation (21). The region involved in the base-pairing interaction is often highly conserved and also referred to as “seed region” (1, 22). Upon base-pairing to target mRNAs, *trans*-acting sRNAs can function as activators or repressors of gene expression by influencing transcript stability and/or translation (17). In Gram-negative bacteria, the base-pairing interaction is frequently mediated by the RNA chaperone Hfq, which will be discussed in detail in section 1.3.1.

The first *trans*-acting sRNA was described in 1984 by *Mizuno et al.* (23). The sRNA named MicF was found to repress the *ompF* mRNA (23). After the first sRNAs were found fortuitously, early systematic searches for sRNAs were based on biocomputational approaches looking mainly for promoter and terminator sequences in intergenic regions as well as homologies and sequences conserved in several species. In combination with experimental approaches, primarily shotgun cloning and microarrays, these studies revealed an increasing number of novel sRNA candidates (24-27). However, by default, these techniques clearly had a strong preference for the identification of sRNAs located in intergenic regions. It is nowadays clear that *trans*-encoded base-pairing sRNAs can also be located in 5' untranslated regions (5' UTRs), 3' untranslated regions (3' UTRs) as well as coding sequences (CDSs), and there are indications that regulatory RNAs even could derive from t-RNA precursors (reviewed in (28)). The advances in RNA high-throughput sequencing (RNA-seq) in the past decade not only led to a constant surge in the number of

## 1 INTRODUCTION

sRNAs identified (29-33), but these new approaches based on RNA-seq also showed that sRNAs are indeed very frequently located in 3' UTRs (reviewed in (34)).

These sRNAs do not necessarily need to have their own promoters, they can also be processed from the 3' UTR by ribonucleolytic cleavage (Fig. 1.1 A). Particularly RNase E, a major ribonuclease in Gram-negative bacteria (discussed in detail in section 1.3.3), has turned out to be a key factor for sRNA biogenesis (35). For the highly conserved sRNA CpxQ for example, which is located in the 3' UTR of the *cpxP* mRNA, *Chao and Vogel* (36) demonstrated that maturation of the sRNA is fully dependent on this ribonuclease. CpxQ has been shown to play an important role during inner membrane stress by repressing multiple mRNA targets (36). In contrast, the *Salmonella* DapZ sRNA is an example for a 3' UTR derived sRNA which expression is differentially regulated by a distinct promoter (Fig. 1.1 B). The sRNA is located in the 3' UTR of the *dapB* mRNA and is transcriptionally activated by HilD, an important regulator of virulence in *Salmonella* (30). Moreover, an sRNA transcribed from its own promoter may also be processed by RNase E into a shorter form (Fig. 1.1 C), as it has been reported for the ArcZ sRNA (35, 37).



**Figure 1.1: sRNA biogenesis from 3' UTRs. A)** Production of the sRNA requires ribonucleolytic cleavage. **B)** The sRNA is transcribed from its own promoter. **C)** The sRNA is transcribed from its own promoter and is processed into a shorter form by ribonucleolytic cleavage.

There are several mechanisms how base-pairing of sRNAs to target mRNAs results in their activation or repression. The classic scenario is an sRNA binding close to the ribosome-binding site (RBS) and preventing association of the 30S ribosomal subunit which results in translation inhibition, as it has been demonstrated for a broad range of sRNAs in a variety of bacteria [e.g. *OxyS-fhlA* in *Escherichia coli* (38), *AcrZ-sdaC* in *Salmonella* (37), *VqmR-vpsT* in *V. cholerae* (39) or *RsaA-mgrS* in *Staphylococcus aureus* (40)]. In the absence of

ribosomes, mRNAs become rapidly degraded by the attack of ribonucleases. However, apart from the canonical binding close to the RBS, there is an increasing number of reported examples where the sRNA base-pairs further upstream in the 5' UTR or further downstream in the coding sequence (reviewed in (41)). Here, other mechanisms than directly blocking ribosome binding come into play. Repression upstream of the RBS can also be achieved by an sRNA binding to translation activator sequences (42-44), whereas repression by an sRNA binding further downstream in the mRNA often involves recruitment of RNase E (45, 46).

Although target repression in general seems to be more frequent, there are several well-studied examples of target activation throughout diverse microbial species (reviewed in (47)). A common mechanism is disruption of an inhibitory structure in the 5' UTR of an mRNA by base-pairing of the sRNA, which otherwise prevents translation of the mRNA (48-50). The first sRNA which was discovered to activate gene expression is the extensively studied RNAIII of the Gram-positive bacterium *S. aureus*. RNAIII activates the *hla* mRNA by binding to an "anti-SD" sequence further upstream in the 5' UTR. In absence of RNAIII, the "anti-SD" binds to the SD sequence, resulting in an inhibitory structure which renders ribosome binding impossible (51). Alternatively, sRNAs can activate targets by sequestering ribonuclease cleavage sites, as it has been reported for example for targets of the *Salmonella* sRNAs SgrS and RydC (52, 53).

Conceptually, there are no differences between the RNA hybrids leading to activation and those leading to repression. It is actually possible that sRNAs act as both activators and repressors on different mRNA targets, even using the very same base-pairing region. In the Qrr sRNAs of *V. cholerae* for example, which are key players to regulate quorum sensing (QS) and virulence in this pathogen, it is the same region that base pairs to the *hapR* mRNA leading to repression and to the *vca0939* mRNA leading to activation (54).

In general, regulation of more than one target by one *trans*-acting sRNA seems to be very common. There are examples where sRNAs regulate several dozens of mRNAs making them global and powerful regulators. Especially in the model organisms *E. coli* and *Salmonella* there are several extensively studied sRNAs, e.g. GcvB (42, 55), RyhB (56, 57), RybB (58, 59) or Spot 42 (60), which control large sets of targets. Further, it is noteworthy that sRNAs can have more than one region used for base-pairing interactions. For example, in the VqmR sRNA of *V. cholerae* three base-pairing regions involved in regulation of different targets have been experimentally confirmed (39, 61). On the other hand, mRNAs can be targeted by more than one sRNA. For instance, the recently identified MicV sRNA in *V. cholerae* shares a large set of targets with the VrrA sRNA, both acting as important regulators in cells experiencing outer membrane stress (62). One of the best studied and most impressive examples for target regulation by multiple sRNAs is the *rpoS* mRNA, which

## 1 INTRODUCTION

is regulated in concert by the sRNAs DsrA, RprA and ArcZ (63-65). The *rpoS* gene encodes the major enterobacterial stress sigma factor  $\sigma^S$  and post-transcriptional activation by these three sRNAs allows the integration of different stress signals in an efficient and elegant manner (reviewed in (66)).

In contrast to all the examples described above, where sRNAs regulate mRNA targets, sRNAs are also able to base-pair to other sRNAs and thereby act as sponges. However, whereas much has been learned about sRNA-mRNA interactions, there are so far only a few well-studied examples of sRNA-sRNA interactions. *Miyakoshi et al.* (67) revealed in *Salmonella* that the conserved GcvB sRNA binds to the SroC sRNA functioning as a sponge and inactivating GcvB. Intriguingly, SroC derives from the polycistronic *gltIJKL* mRNA, which is targeted by GcvB. Binding of SroC to GcvB results in derepression of the GcvB mRNA targets. Recently, *Melamed et al.* (68) found that the *E. coli* RbsZ sRNA base-pairs to and downregulates the extensively studied RybB sRNA. Although the physiological role of the sponging interaction remains elusive, it establishes an interesting autoregulatory loop, as RbsZ corresponds to the 3' end of the *rbsB* mRNA, which is targeted by RybB.

### 1.2.2 Physiological roles of sRNAs and their impact on regulatory networks

Given that sRNAs are extremely versatile regulators being able to control a large range of targets, it is not surprising that sRNAs are involved in a broad variety of bacterial processes. They are often expressed under specific conditions, allowing the organism to rapidly adjust gene expression when facing particular environmental challenges. Typically, sRNAs are entangled in modulating metabolic pathways and responding to nutrient limitations and diverse stress factors enabling bacteria to survive even under harsh conditions (69, 70). In pathogenic bacteria, sRNAs are often important players in regulating genes required for colonization of the host and virulence (71). Furthermore, bacterial sRNAs have been reported to be involved in cell-to-cell communication (quorum sensing) and biofilm formation in various species (72).

As powerful regulators in these diverse physiological processes, sRNAs are very often part of complex regulatory circuits and large networks. The rising number of sRNA-target connections identified in the last two decades revealed that typically recurring motifs, such as feed forward loops or positive and negative feedback regulation, originally described for transcriptional regulatory circuits, frequently involve both sRNAs and transcription factors forming so called mixed regulatory networks (reviewed in (73-75)).

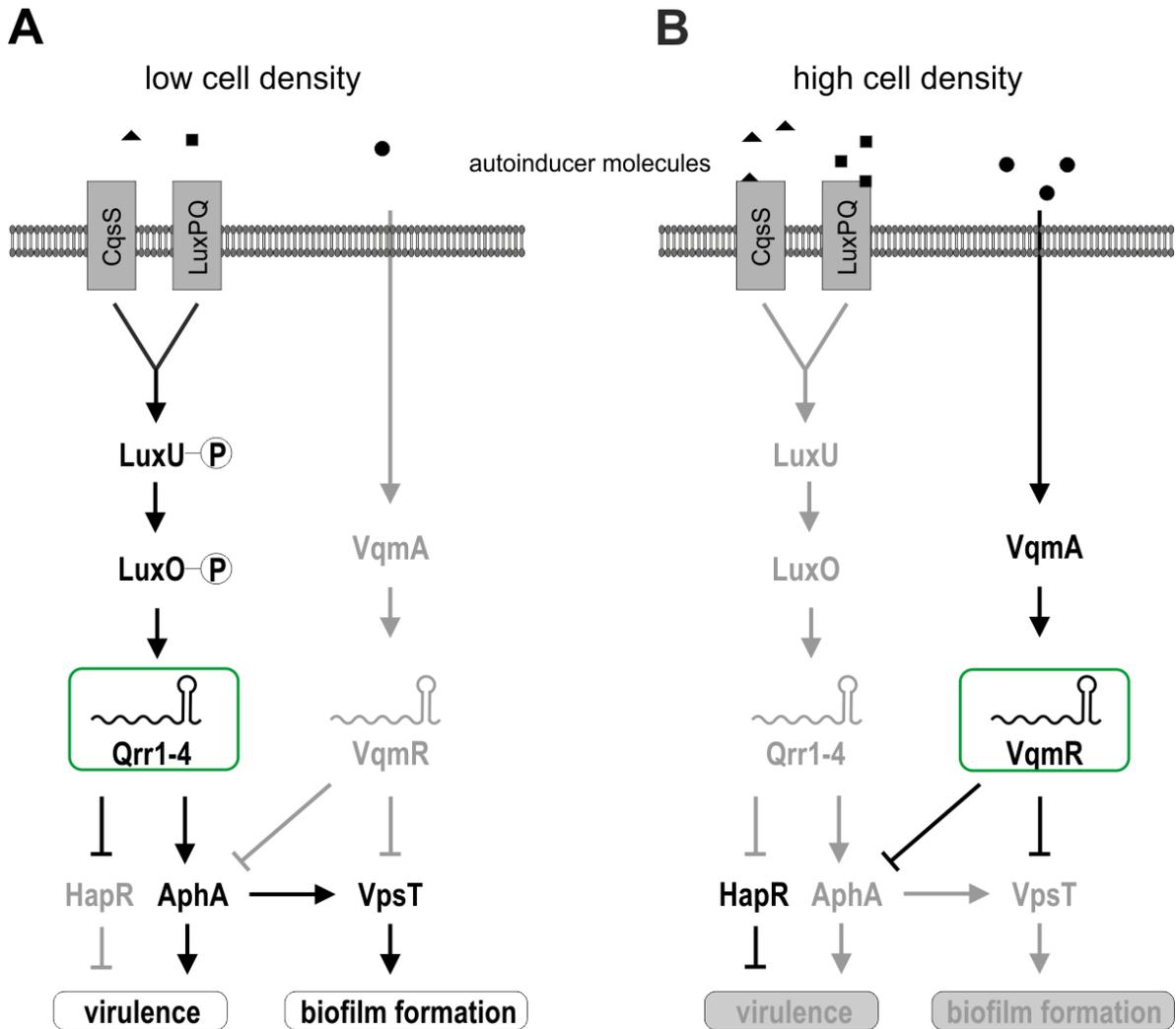
The regulon of the highly conserved Spot 42 sRNA is a well-studied example where an sRNA regulates a very large set of targets to adjust metabolism in response to the availability of different carbon sources. Spot 42, encoded by the *spf* gene, is repressed by cAMP-CRP (cAMP receptor protein) in the absence of glucose, the preferred carbon source of *E. coli* cells (76). Under glucose-rich conditions, Spot 42 is strongly expressed and post-transcriptionally represses numerous genes involved in the uptake and catabolism of non-preferred carbon sources, many of which are transcriptionally activated by cAMP-CRP (60). This establishes a multi-output coherent feed forward loop in which the Spot 42 sRNA plays a pivotal role to dynamically and efficiently regulate catabolite repression in *E. coli* (60, 77). In the last years, an increasing number of additional Spot 42 targets have been proposed by different biocomputational and experimental approaches, further underlining the global impact of Spot 42 (reviewed in (78)).

The quorum sensing (QS) circuit of the *Vibrionaceae* is another paradigm of sRNA-based regulation. QS is a cell-to-cell communication process based on the production, release and detection of extracellular signaling molecules, which allows bacteria to regulate their gene expression in response to cell population density and thereby coordinate group behavior (79). In *V. cholerae*, four homologous sRNAs, the quorum regulatory RNAs Qrr1-4, act at the center of the QS pathway (80). They are expressed at low cell density (LCD), when the concentration of signaling molecules, called autoinducers, is low. Phosphate is funneled from the membrane-bound receptors via the phosphor-transfer protein LuxU to the response regulator LuxO, which, in its phosphorylated form, activates transcription of the Qrr sRNAs (80, 81). The Qrrs post-transcriptionally activate the LCD master regulator AphA and repress the high cell density (HCD) master regulator HapR, two transcription factors controlling antagonistically virulence and biofilm formation (82, 83) (Fig. 1.2 A). In addition, expression of *aphA* is regulated by another, independent QS pathway controlled by the VqmR sRNA. VqmR post-transcriptionally represses *aphA* as well as other targets including *vpsT*, encoding a major activator of biofilm formation, at HCD (61) (Fig. 1.2 B). Together, Qrr1-4 and VqmR are key players orchestrating virulence and biofilm formation in *V. cholerae*. Several negative feedback loops are involved in the signal transduction through the QS circuit. *Svenningsen et al.* (84) showed that the HapR protein activates transcription of the *qrr* genes, whereas the Qrr sRNAs post-transcriptionally repress the *hapR* mRNA. Furthermore, a negative feedback loop between LuxO and the Qrrs has been demonstrated (85). These two feedback loops allow a gene dosage compensation of the four redundant Qrr sRNAs and contribute to a tight control of QS (85).

The prevalence of sRNAs in diverse networks and physiological processes revealed in the last years raises the question of what are the advantages and differences of sRNA-based regulation compared to protein-based transcriptional regulation. Several aspects

## 1 INTRODUCTION

should be considered (73-75, 86). First, with sRNA-based regulation occurring at the post-transcriptional level, the desired regulatory effect might be reached faster, which can be crucial especially when bacteria encounter sudden changes in the environment. For instance, unwanted protein synthesis can be blocked immediately by a base-pairing sRNA inhibiting translation. This means, mRNAs which are already made can be repressed, whereas transcription factors can only act one level above. Further, sRNAs can substantially expand a regulatory network and increase possible regulatory outcomes. By binding of several sRNAs to one mRNA, for example, different signals can be integrated in order to fine-tune gene expression. Base-pairing sRNAs also allow a discoordinate regulation of genes transcribed from one operon. Another key aspect are the different regulatory dynamics between sRNAs and transcription factors. Whereas transcription factors act catalytically, sRNAs are often degraded together with their targets (coupled degradation). However, catalytic degradation of the target or pure sequestration are also possible outcomes (87, 88). When sRNAs act in a stoichiometric manner, the ratio between sRNA and mRNA molecules influences the regulatory output. Consequently, the abundance of one target mRNA can influence the expression of another one, establishing a cross talk between different targets and circuits. Apart from these, further intriguing roles of sRNAs in regulatory networks have been described and continue to be discovered (reviewed in (73-75, 86)). Studying bacterial regulatory networks is a burgeoning field and increasing knowledge of global sRNA-target interactions contributes to a deeper understanding of the interplay of different regulators in the cell.



**Figure 1.2: Quorum sensing in *V. cholerae*.** Active factors are shown in black, inactive factors are shown in gray. **A)** At low cell density, autoinducer concentrations are low, membrane bound receptors (CqsS and LuxPQ) act as kinases, phosphate is channeled from LuxU to LuxO, and transcription of the *Qrr* sRNAs (*Qrr1-4*) is activated. *Qrr1-4* post-transcriptionally repress *hapR* and activate *aphA*, resulting in expression of virulence and biofilm genes. **B)** At high cell density, binding of autoinducer molecules converts the receptors to phosphatases. Consequently, *Qrr1-4* are not expressed, HapR is produced. The transcription factor VqmA, belonging to an independent QS pathway, activates expression of the *VqmR* sRNA, post-transcriptionally repressing *aphA* and *vpsT*.

## 1.3 Proteins required for regulation by sRNAs

### 1.3.1 The RNA chaperone Hfq

Bacterial sRNAs most often do not act in isolation, but require the aid of RNA chaperones to carry out their function. One major bacterial RNA chaperone is Hfq. This RNA-binding protein was originally identified in the 1960s as an essential host factor for phage  $\text{Q}\beta$  replication in *E. coli* (89). Since then, the Hfq protein has attained enormous attention and has been subject of countless studies. Hfq is widely distributed. It is assumed that at least 50% of all bacterial species encode an Hfq homolog (90, 91), and distant homologs are also found in archaea and eukaryotes, indicating an ancient origin of the protein (91-93). Belonging to the Sm and Sm-like (Lsm) protein family, Hfq is characterized by a conserved sequence motif forming a stable structural element, the Sm domain (94). In contrast, the C-terminus of the protein is intrinsically disordered and poorly conserved (92). The Hfq protomers assemble into a homohexamer and form a ring-like structure (94, 95).

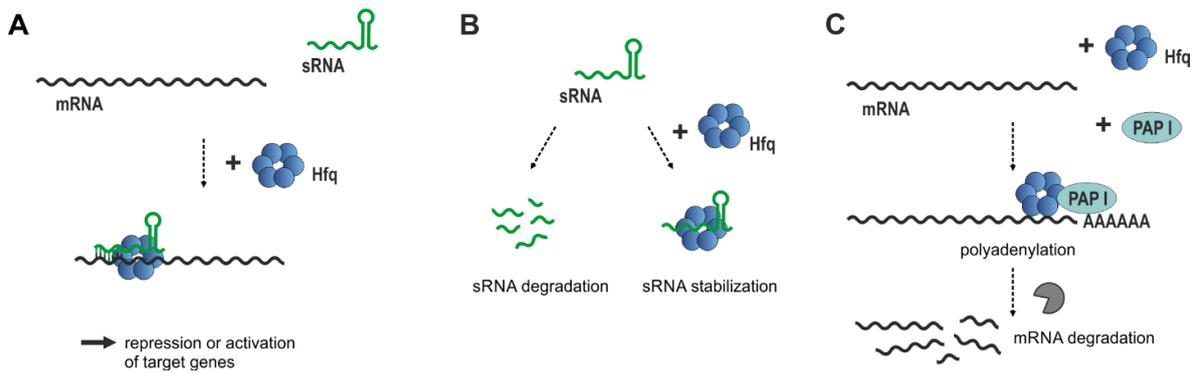
Hfq can interact with RNAs at four different sites of the hexameric ring. They are commonly referred to as proximal site, distal site, rim and C-terminal tail, and play different roles for RNA binding. As RNA-ligands of Hfq vary tremendously in size, structure and sequence, much effort has gone into studying how Hfq recognizes and discriminates between its substrates. It is now well established that sRNAs primarily bind to the proximal face of the ring, with the Rho-independent terminator of many sRNAs, particularly the poly(U)-tail, being a crucial feature for recognition (96, 97). However, it is not exclusively the proximal site which is relevant for sRNA binding. *E. coli* sRNAs have been divided into two groups based on how they bind Hfq: sRNAs binding to the proximal face and the rim are called class I sRNAs, whereas sRNAs binding to the proximal and the distal face are referred to as class II sRNAs (98). In contrast to the proximal face, the distal face of Hfq preferentially binds to A-rich sequences, often found in the 5' UTR of mRNAs. An A-R-N motif was proposed by *Link et al.* (99), and has been refined by other studies (100-103). Moreover, it has also been shown that the position of the A-rich motif relative to the base-pairing region plays a critical role (104, 105). The outer rim of the hexamer and the C-terminal domain seem to have additional functions for RNA binding. The rim binds to UA-rich sequences, allowing further stabilizing contacts (106-108), and thereby facilitating the annealing interaction between two RNA molecules (109). The role of the C-terminus has been subject of controversial discussions for a long time. Increasing evidence, however, suggests that the C-terminal tail is also critical at least for some sRNAs (108, 110), and seems to be important for the release of the RNA duplexes formed on Hfq (111).

Based on these diverse binding capacities, Hfq's main function is to facilitate the base pairing interaction between an sRNA and an mRNA. Hfq acts as a "molecular matchmaker" helping sRNAs to find, base-pair to and regulate their targets (92). In the currently established model, an sRNA and an mRNA are simultaneously bound to Hfq and a transient tripartite Hfq-sRNA-mRNA complex is formed, facilitating the duplex formation between the sRNA and the mRNA (92) (Fig. 1.3 A). The Hfq-mediated binding of an sRNA to its target can result in activation or repression of the target (see section 1.2.1). For the latter, Hfq, together with the bound sRNA, can also actively recruit RNase E (112, 113) (described in detail in section 1.3.3). However, Hfq's function is not limited to this. The protein also plays a key role in stabilizing sRNAs by protecting them from ribonucleolytic cleavage (114) (Fig. 1.3 B). Reduced lifetimes in the absence of Hfq have been reported for several sRNAs in various bacteria (30, 39, 62, 115). Further, it has been demonstrated that Hfq can promote the polyadenylation of mRNAs by polyA-polymerase (PAP I), leading to 3' to 5' degradation by exoribonucleases (116, 117) (Fig. 1.3 C). Moreover, it has been revealed that Hfq alone is able to inhibit mRNA translation by binding close to the ribosome binding site (118).

Nevertheless, the main function of Hfq is that of a "molecular matchmaker" (119, 120), and several studies have been performed to better understand the global role Hfq plays in the cell and the underlying mechanistic details. It has been shown that Hfq is a limiting factor in the cell, and that RNA molecules are typically competing for access to the protein (121). According to *Fender et al.* (122), RNA molecules are continuously cycling on and off Hfq. In contrast to the dynamic target regulation by Hfq-binding sRNAs *in vivo*, Hfq-RNA complexes show very low dissociation rates *in vitro*. This apparent conflict was solved by the "active cycling" model, in which competitor RNAs transiently bind to Hfq and thereby speed up the dissociation of resident RNAs (122, 123). The fact that there is a competition between RNA molecules for binding to Hfq implicates that different regulatory networks are not only connected through the sRNAs themselves, but also through the availability of Hfq (124). Moreover, competition for Hfq does not only involve sRNAs, but also mRNA targets. Competition between targets has indeed been demonstrated to be a major factor for regulatory outcomes (125).

Although there are still unresolved questions about how exactly Hfq finds its targets and is able to discriminate between the thousands of RNAs in the cell, there is no doubt of the immense and global impact Hfq has on bacterial post-transcriptional gene regulation. Given that Hfq binds hundreds of different RNAs (126), it is not surprising that *hfq* mutants show severe and pleiotropic phenotypes (127). Decreased growth rates, reduced stress tolerance and loss of virulence have been reported for a broad spectrum of bacteria (128-131).

## 1 INTRODUCTION



**Figure 1.3: Different functions of Hfq.** **A)** Hfq mediates the base-pairing interaction between an sRNA and its target mRNA. This can result in repression or activation of the target. **B)** Hfq stabilizes sRNAs by protecting them from ribonucleolytic cleavage. **C)** Hfq promotes polyadenylation of mRNAs by polyA-polymerase (PAP I), leading to degradation by 3' to 5' exonucleases.

### 1.3.2 Other RNA-binding proteins

Although Hfq affects a very large set of RNAs in the cell, it is not the only RNA-binding protein (RBP) involved in bacterial post-transcriptional gene regulation. Another widespread and well-studied RBP is CsrA, originally identified in *E. coli* as a regulator of glycogen biosynthesis (132). In contrast to Hfq, CsrA does not act as a molecular matchmaker. By binding to multiple GGA motifs, the protein usually binds to the 5' UTR of coding sequences, altering their translation. The activity of CsrA is modulated by decoy sRNAs containing numerous GGA recognition sites and being able to bind several CsrA molecules at the same time (reviewed in (18, 133), see also section 1.1).

Apart from the two well established RBPs Hfq and CsrA, the role and function of a third globally acting sRNA-related RBP is just emerging. ProQ, a FinO-domain containing protein and originally described as a regulator of the *E. coli* proline transporter ProP (134), has recently been shown to interact with hundreds of RNAs in *E. coli* and *Salmonella* and to govern a large post-transcriptional network similar to Hfq (68, 135, 136). Recent evidence suggests that ProQ recognizes its targets primarily by structure. Using CLIP-seq analysis, no specific sequence motif could be identified; however, ProQ binding sites were mostly detected close to or within strong secondary structures (136). Similar to Hfq, ProQ seems to stabilize sRNAs and mediate duplex formation between sRNAs and mRNAs (135, 137). However, in contrast to Hfq, many interactions with ProQ were found in mRNA 3' ends, suggesting additionally an important role in stabilizing mRNAs by protecting them from 3' exonucleolytic cleavage (136). Interestingly, *Melamed et al.* (68) revealed that ProQ and Hfq

have a relevant fraction of shared targets whereas the majority of ligands associate either with ProQ or Hfq. This is in line with a previous study suggesting that ProQ preferentially binds to Hfq-independent sRNAs (135). Understanding the overall function of ProQ in the cell will require further research, but undoubtedly, it seems to be another key player in bacterial post-transcriptional gene regulation.

Beside such universal RNA-binding proteins, others with more specialized functions have been described, too. For instance, the *E. coli* CspA and related cold shock proteins bind to RNA and unfold secondary structure when cells are experiencing low temperature (138). RapZ (RNase adaptor protein for sRNA GlmZ) is an example for an RNA-binding protein with a very specific function: Its only binding partners are the sRNAs GlmZ and GlmY. The protein here acts as an adapter to guide RNase E to GlmZ, whereas GlmY acts to sequester RapZ (139).

### 1.3.3 The ribonuclease RNase E

Not only RNA-binding proteins are required for successful regulation of gene expression by sRNAs. Another key player is RNase E, a ribonuclease which is conserved in gamma-proteobacteria and which has homologs in many other bacterial species (140). RNase E is a large protein, acting as a homotetramer, and consisting of an N-terminal catalytic domain and a C-terminal non-catalytic domain (141, 142). The C-terminal domain is natively unstructured and serves as a scaffold for the assembly with other factors (142). In *E. coli* and related bacteria, these include polynucleotide phosphorylase (PNPase), an RNA helicase (RhlB) and the glycolytic enzyme enolase. Together, RNase E and these proteins, can form a multiprotein complex referred to as the degradosome (143). Degradosome-like complexes seem to be present in many proteobacteria, however, the exact composition varies (144).

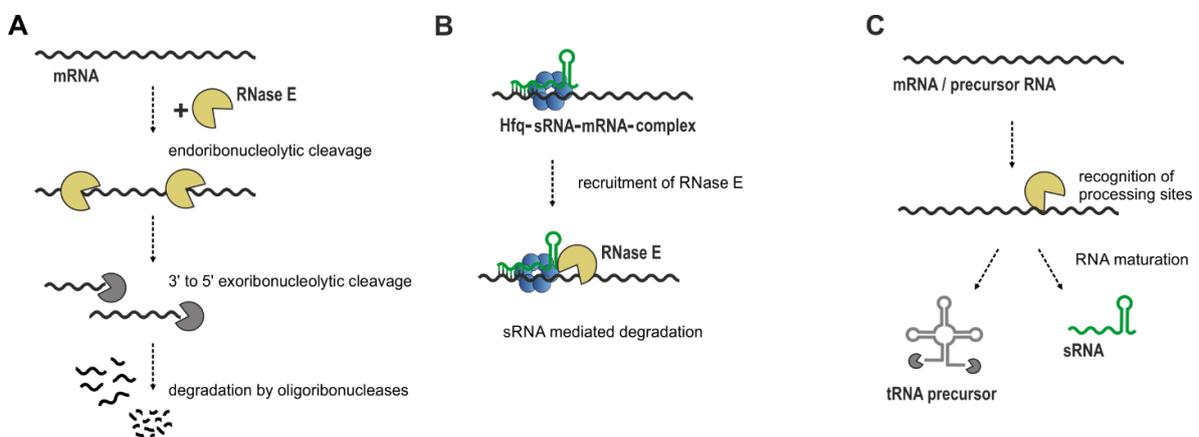
In general, RNA degradation, which might primarily be perceived as a process to get rid of RNA molecules not needed anymore, has a pivotal role in regulating gene expression, as it controls steady-state levels of mRNAs to be translated. Degradation of mRNA molecules is usually initiated by endonucleolytic cleavage. In Gram-negative bacteria, RNase E has been revealed as a major player here, but other endonucleases, such as RNase G or RNase III might be involved as well. After transcripts are attacked by these endonucleases, resulting RNA fragments with new unprotected 3' ends are degraded by 3' to 5' exoribonucleases. Subsequently, oligoribonucleases are required for further degradation of the remaining short fragments into mononucleotides (145) (Fig. 1.4 A).

## 1 INTRODUCTION

Apart from its function in initiating general RNA turnover, RNase E plays a pivotal role in sRNA-mediated degradation of target transcripts (112, 113, 115) (Fig. 1.4 B). This was shown, for example, for the Hfq-binding sRNA RyhB, which is degraded together with its targets in an RNase E dependent process, referred to as coupled degradation (115). *Morita et al.* (112) demonstrated that RNase E can be co-purified together with Hfq and sRNAs, indicating the formation of a ribonucleoprotein complex, specialized in sRNA-mediated target degradation and clearly different from the degradosome complex.

RNase E is not only a key factor for RNA degradation, but also for RNA processing (Fig. 1.4 C). Beside its role in rRNA and tRNA maturation (146-148), RNase E has recently turned out to be a central player for sRNA biogenesis as well (35, 149). By applying the TIER-seq (transiently inactivating an endoribonuclease followed by RNA-seq) approach, *Chao et al.* (35) identified RNase E cleavage sites in *Salmonella* at a global scale and showed that a large number of 3' UTR derived sRNAs require RNase E for expression (see also section 1.2.1).

The various functions of RNase E bring up the question of how RNase E recognizes its substrates. RNase E is a single-strand-specific RNase with a preference for AU-rich regions (150, 151). Recently, *Chao et al.* (35) reported that a uridine at position +2 is a key feature for cleavage. Regarding substrate recognition by RNase E, two different pathways have been described: a 5' end dependent and an internal direct entry pathway (152). In the first one, which relies on a 5' sensor within the N-terminal domain, 5' monophosphate is strongly preferred over 5' triphosphate (141, 153). The second pathway is 5' end independent (154, 155), which means that RNase E here bypasses the RNA 5' end. Recent evidence suggests that for this direct entry of RNase E, stem loop structures might play an important role for recognition (156). It seems likely that the two different modes of action of RNase E do not exclude each other but are combined to ensure optimal recognition and activity (156).



(figure legend next page)

**Figure 1.4: Different functions of RNase E.** **A)** Endoribonucleolytic cleavage by RNase E results in RNA fragments that are degraded by 3' to 5' exoribonucleases. Remaining RNA degradation products are subsequently digested by oligoribonucleases. Other components of the degradosome RNase E can assemble with are not shown for the sake of simplicity. **B)** RNase E is involved in sRNA-mediated target degradation. Here, RNase E can be actively recruited by an Hfq-sRNA-complex. **C)** RNase E is involved in the processing and maturation of different types of RNAs.

## 1.4 Characterization and identification of Hfq-binding sRNAs

Given the various sRNAs which have been reported to associate with Hfq, the question arises of what are their common features and what distinguishes them from other regulatory RNAs. Based on extensive studies in different organisms, it seems that Hfq-binding sRNAs usually have a length between 50 and 250 nucleotides and often contain hairpin structures (92). Being a diverse group lacking a specific sequence motif, Hfq-binding sRNAs can rather be characterized by their “modular” organization involving three common features (1, 92): (i) a 3' end stem-loop followed by a poly(U)-tail promoting Rho-independent transcription termination, (ii) a binding site for Hfq, often close to the terminator, and (iii) a sequence stretch base-pairing to targets. The base-pairing region is usually highly conserved and frequently, but not always, located close to the 5' end. The modular character of the base-pairing region was in particular highlighted by *Papenfort et al.* (157) showing that this domain can still fulfill its regulatory function, even when fused to another sRNA scaffold. In this study, the conserved seed region of the RybB sRNA was combined with an unrelated sRNA backbone and the resulting chimeric sRNA still repressed native RybB targets.

Due to these modular characteristics, Hfq-binding sRNAs are often compared to eukaryotic microRNAs (22), frequently considered as their functional equivalents. Although the latter are shorter and, in contrast to bacterial sRNAs, with a defined length of 21 to 23 nucleotides, there are clear parallels in the capacity of binding to multiple RNAs via a seed region and acting post-transcriptionally to repress gene expression (22, 158). A specific characteristic of microRNAs, however, is their association with Argonaute (AGO) proteins within an RNA-induced silencing complex (RISC) required for the microRNA-based gene regulation (159).

For both microRNAs and bacterial sRNAs, bioinformatic as well as experimental high-throughput based approaches have been used to predict ever more novel candidates in the last two decades (32, 39, 160-163). Studying these in conjunction with the proteins they are binding to (AGO proteins in case of microRNAs and Hfq or other RNA-binding proteins in case of bacterial sRNAs) has turned out to be very helpful here, and several similar methods,

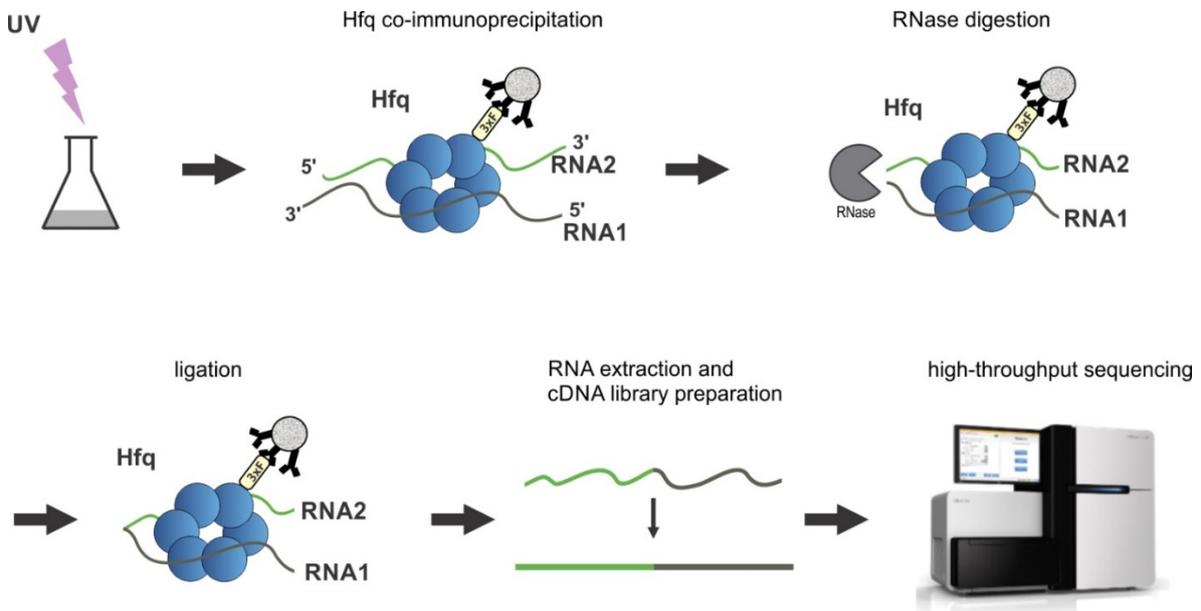
## 1 INTRODUCTION

usually based on co-immunoprecipitation of RNA with the protein of interest, have been developed in parallel in eukaryotic and bacterial cells for this purpose.

In this context, the established method to identify Hfq ligands in bacterial cells is RNA co-immunoprecipitation followed by RNA-sequencing (RIP-seq) (reviewed in (3)). RIP-seq is based on the pull-down of the desired protein together with its RNA-ligands, followed by RNA extraction, cDNA library preparation and high-throughput sequencing. One pioneering study here was performed in 2008 by *Sittka et al.* (29), where hundreds of mRNAs and several dozens of sRNAs were predicted to associate with Hfq in *Salmonella*, and it was shown that with this method, Hfq-binding sRNAs can be recovered with high specificity. Since then, the approach has been successfully applied to other bacterial species, as for example *Neisseria meningitidis* (31). Having information about binding of an sRNA to a specific RBP can provide important indications for its function in the cell, and help to distinguish transcriptional noise and degradation products from functional transcripts (86). But RIP-seq not only sheds light on Hfq association of previously suggested sRNA candidates, it also allows the identification of hitherto completely unknown sRNAs, including many deriving from 3' UTRs by ribonucleolytic cleavage (30), which had not been detected by other approaches before. Furthermore, the enrichment by the co-immunoprecipitation procedure facilitates the discovery of poorly expressed sRNAs.

Based on RIP-seq, several further techniques have been developed, which include additional steps or variations in order to obtain more detailed information. For instance, in the CLIP-seq method (crosslinking and immunoprecipitation followed by RNA-sequencing), a UV-crosslinking step is introduced before the co-immunoprecipitation procedure to covalently link bound RNAs to Hfq. Mutations induced by the crosslinking reveal further information about the binding site of Hfq (164). This is analogous to HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation) in eukaryotic cells, employed to identify interaction sites of microRNAs with for example the AGO proteins (165-167). Another recently developed method is RIL-seq (RNA interaction by ligation and sequencing), which combines co-immunoprecipitation with the ligation of RNA molecules simultaneously bound to Hfq (126, 168) (Fig. 1.5). Apart from identifying Hfq binding partners, this approach allows to detect the direct interaction of two RNA molecules, providing additional data about Hfq mediated target regulation. Similar to RIL-seq is the CLASH (cross-linking, ligation and sequencing of hybrids) technique, developed in eukaryotic cells in order to identify microRNA targets (169), but also adapted to bacteria (170, 171). The various high-throughput based methods developed in the last ten to twenty years (reviewed in (3, 172)) are undoubtedly powerful tools to study RNA-RNA and protein-RNA interactions in bacterial cells and to get deeper insights into RNA-based regulatory

mechanisms on a global scale. However, one major future challenge will be to combine, compare and validate the accumulating huge data sets.



**Figure 1.5: Visualization of the RIL-seq method.** After *in vivo* RNA-protein cross-linking, cells are lysed and subjected to co-immunoprecipitation. Exposed regions of bound RNAs are trimmed by RNases, and RNA molecules are ligated. Subsequently, RNA is extracted, cDNA libraries are prepared and subjected to high-throughput sequencing. cDNA reads may then derive from two distinct RNA molecules which were simultaneously bound to Hfq.

## 1.5 Model organisms used in this study

In this thesis, different model organisms are used. Chapters 2, 3 and 5 present studies on *V. cholerae* sRNAs. *V. cholerae* is a major human pathogen and an important model organism for pathogenic bacteria. Chapter 4 presents one of the few studies on sRNAs in *Photorhabdus laumondii* and *Xenorhabdus szentirmaii*, two entomopathogenic species mainly known for secondary metabolite production.

### 1.5.1 *Vibrio cholerae*

*V. cholerae* is a Gram-negative, comma-shaped bacterium belonging to the *Vibrionaceae*, a family within the gamma-proteobacteria. As a facultative anaerob, its natural habitat is in the water, often in estuarine ecosystems (173). The bacterium is a human pathogen, causing the diarrheal disease cholera. Several serotypes exist, of which two (serotypes O1 and O139) are important for epidemics (174). The current and seventh cholera pandemic is caused by *V. cholerae* O1 El Tor biotype (174), which is also used as model organism in this study. Its genome, distributed on two circular chromosomes, is fully sequenced (175).

Characteristic for *V. cholerae*'s lifestyle is the ability to switch between planktonic behavior and biofilm formation. *V. cholerae* has a polar flagellum, allowing the bacterium to be highly motile in the planktonic state (176). As many other bacteria, *V. cholerae* can form biofilms, increasing the resistance against diverse stresses (177). Biofilms are communities of bacterial cells attached to a surface and to each other, and enclosed in an extracellular matrix, consisting mainly of polymeric substances produced by the cells in the biofilm. Importantly, cells in a biofilm also show altered gene expression (178, 179).

*V. cholerae* forms biofilms in both the aquatic environment and during infection of the human host. In aquatic ecosystems, where biofilms can protect, for example, against predators and antimicrobial substances, *V. cholerae* preferentially forms biofilms on chitinous surfaces, e.g. zooplankton (177). Chitin can be used as a carbon source (180), and induces competence, allowing the uptake of new genetic material by natural transformation (181). However, several studies showed that biofilm formation also plays a crucial role during different stages of infection and transmission of *V. cholerae* cells (reviewed in (182)). For instance, embedment in a biofilm seems beneficial to survive the acidic barrier of the stomach cells have to pass after entering the host (183).

Once *V. cholerae* cells are in the human host, two major virulence factors are important for their pathogenicity: The toxin-coregulated pilus (TCP), encoded on the *Vibrio* pathogenicity island (VPI), is required for colonization of the intestine (184-186). *Vibrio* then produces cholera toxin (CT), a protein causing severe watery diarrhea leading to strong dehydration and, if untreated, death (187). The cholera disease is still a burden in underdeveloped countries with poor hygiene conditions (174, 188). It is caused by the uptake of contaminated food or water (187), leading to about 4 million cases a year, according to the World Health Organization.

In the human gut as well as in aquatic habitats, *Vibrio* constantly faces environmental changes. In particular, the transition between the host and the natural habitat is accompanied by a number of sudden alterations. To cope with these, *Vibrio* needs to quickly adapt and regulate gene expression accordingly. Despite the pivotal role of sRNAs in

guaranteeing dynamic regulation of gene expression that has been shown in the last decades, still relatively little is known about the regulatory roles of sRNAs in *V. cholerae*. Apart from the sRNAs involved in quorum sensing (see section 1.2.2), where *Vibrio* species served as important model organisms (80, 82), there are only relatively few sRNAs which have been studied in detail (189, 190). Examples are the  $\sigma^E$  dependent sRNAs MicV and VrrA, regulating the expression of several outer membrane proteins (62, 191), or the TarA and TarB sRNAs, establishing an important link between the virulence regulatory pathway and glucose uptake (192, 193). Some other sRNAs which are highly conserved and have been extensively described in the model organisms *E. coli* and *Salmonella*, have been found in *V. cholerae*, too, due to the conservation of sequence or function, e.g. GcvB (194) and RyhB (195, 196). However, although dozens of sRNAs have been predicted in *V. cholerae* by high-throughput sequencing based approaches (39, 197), the number of sRNAs studied comprehensively so far remains quite limited. Moreover, whereas RNA-ligands of Hfq and other RNA-binding proteins have been studied on a global scale in other bacteria (29, 31, 136, 164, 198), in *V. cholerae* so far information about association with Hfq, CsrA or ProQ has been lacking.

### 1.5.2 *Photorhabdus laumondii* and *Xenorhabdus szentirmaii*

*P. laumondii* and the closely related *X. szentirmaii* are Gram-negative bacteria of the family *Enterobacteriaceae*, belonging, as the *Vibrionaceae*, to the gamma-proteobacteria. The two species live in a symbiotic relationship with entomopathogenic nematodes of the genera *Heterorhabditis* and *Steinernema*, respectively (199, 200). In a complex life cycle, the bacteria and the nematodes cooperate to colonize and kill the insect host. *P. laumondii* and *X. szentirmaii* first reside in the gut of the juvenile stage of the soil-dwelling nematodes. When the nematodes infect an insect, they migrate to the haemolymph and release the bacteria. The bacteria then kill the insect by the production of toxins. Subsequently, the insect cadaver serves as source of nutrition for both the bacteria and the nematode, and finally the bacteria start a new mutualistic association with the next generation of nematodes. Development of the nematodes is dependent on the symbiosis with the bacteria (199, 200).

A characteristic trait of *P. laumondii* and *X. szentirmaii* is the production of specialized metabolites. Several biosynthesis gene clusters encoding these secondary metabolites have been identified in both species (201). They comprise a number of structurally different classes, as e.g. peptide antibiotics (*P. laumondii* and *X. szentirmaii*), anthraquinones (*P. laumondii*) or pyrrothine metabolites (*X. szentirmaii*). Not only have these specialized metabolites an ecological function and are important for the symbiotic lifecycle; they are also

of industrial relevance and can serve as a potential source for the development of new pharmaceuticals (201, 202).

It has been shown previously that Hfq is a key factor for the production of specialized metabolites (203). Moreover, *hfq* mutants of *P. laumondii* and *X. szentirmaii* are no longer able to successfully maintain their symbiosis with nematodes (203). However, the binding partners of the RNA chaperone had not been identified, and the exact mechanism remained elusive. Generally, knowledge about small regulatory RNAs is so far very limited in these two species.

### 1.6 Aim and scope of the study

The RNA chaperone Hfq has undoubtedly been established as a widespread key factor for post-transcriptional gene regulation in bacteria. It has been shown that a vast majority of *trans*-acting sRNAs require Hfq to fulfill their regulatory function (28). Nevertheless, despite the global role Hfq-binding sRNAs seem to play for the physiology of many bacteria (127), detailed knowledge about the RNA-ligands of Hfq is limited to rather few organisms, and most of our knowledge about sRNAs comes from studies with the model organisms *E. coli* and *Salmonella*.

The overall aim of this thesis is to identify and characterize Hfq-binding sRNAs with a focus on the human pathogen *V. cholerae*, as well as in the entomopathogenic bacteria *P. laumondii* and *X. szentirmaii*. One goal is to investigate specific representative sRNAs in detail, focusing on their biogenesis, structure, base-pairing with targets and association with RNA-binding proteins. Furthermore, global approaches are applied to identify Hfq-binding sRNAs and their targets on a large scale in order to gain deeper insights into the regulatory networks governed by Hfq.

In the first project (chapter 2), we address the question of which RNAs are bound to Hfq in *V. cholerae*. To investigate this from a global perspective, we use RIP-seq analysis, and subsequently focus on Hfq-binding sRNAs identified here. We investigate one representative example in more detail, analyzing its expression, Hfq dependency and biogenesis. Moreover, we study how the sRNA, which we name FarS, regulates target mRNAs, establishes a feed forward regulatory loop and thereby acts as an important player in the fatty acid metabolism of *V. cholerae*.

In the second project (chapter 3), we aim to further explore sRNAs processed from 3' UTRs of coding sequences. To this end, RNase E cleavage sites in *V. cholerae* are determined genome-wide by TIER-seq. We then focus on two Hfq-binding sRNAs, OppZ

and CarZ, which, interestingly, base-pair with and repress the respective mRNA transcripts they are clipped-off, raising the question if they could have an autoregulatory function. Subsequently, we investigate how these sRNAs establish a negative feedback loop at the post-transcriptional level.

In the project in chapter 4, our goal is to identify Hfq ligands in *P. laumondii*, and to identify sRNAs in this model organism as well as in the very closely related *X. szentirmaii*, two species in which sRNAs have been rarely studied. Moreover, we examine the function of the conserved ArcZ sRNA and the previously unknown role it plays in regulating production of specialized metabolites in *P. laumondii*.

In the last project (chapter 5), we apply the RIL-seq method to *V. cholerae* Hfq in order to globally identify RNA-RNA interactions mediated by this chaperone and thereby facilitate the challenging task of revealing sRNAs targets. We examine this large scale data set in detail, and then focus on a novel sRNA base-pairing with the quorum sensing regulatory sRNAs (Qrr1-4). We investigate how this sRNA, which we name QrrS (**Qrr1-4 sponge**), acts as a hitherto unknown sponge RNA modulating quorum sensing dynamics in *V. cholerae*.



## 2 Switching fatty acid metabolism by an RNA-controlled feed forward loop

**Michaela Huber**, Kathrin S. Fröhlich, Jessica Radmer, and Kai Papenfort (2020).  
“Switching fatty acid metabolism by an RNA-controlled feed forward loop.” *Proc Natl  
Acad Sci U S A* 117(14): 8044-8054.

[doi.org/10.1073/pnas.1920753117](https://doi.org/10.1073/pnas.1920753117)





# Switching fatty acid metabolism by an RNA-controlled feed forward loop

Michaela Huber<sup>a,b</sup>, Kathrin S. Fröhlich<sup>a,b</sup>, Jessica Radmer<sup>b</sup>, and Kai Papenfort<sup>a,b,c,1</sup>

<sup>a</sup>Institute of Microbiology, Friedrich Schiller University, 07745 Jena, Germany; <sup>b</sup>Faculty of Biology I, Ludwig-Maximilians-University of Munich, 82152 Martinsried, Germany; and <sup>c</sup>Microverse Cluster, Friedrich Schiller University Jena, 07743 Jena, Germany

Edited by Susan Gottesman, National Institutes of Health, Bethesda, MD, and approved February 19, 2020 (received for review November 26, 2019)

**Hfq (host factor for phage Q beta) is key for posttranscriptional gene regulation in many bacteria. Hfq's function is to stabilize sRNAs and to facilitate base-pairing with *trans*-encoded target mRNAs. Loss of Hfq typically results in pleiotropic phenotypes, and, in the major human pathogen *Vibrio cholerae*, Hfq inactivation has been linked to reduced virulence, failure to produce biofilms, and impaired intercellular communication. However, the RNA ligands of Hfq in *V. cholerae* are currently unknown. Here, we used RIP-seq (RNA immunoprecipitation followed by high-throughput sequencing) analysis to identify Hfq-bound RNAs in *V. cholerae*. Our work revealed 603 coding and 85 noncoding transcripts associated with Hfq, including 44 sRNAs originating from the 3' end of mRNAs. Detailed investigation of one of these latter transcripts, named FarS (fatty acid regulated sRNA), showed that this sRNA is produced by RNase E-mediated maturation of the *fabB* 3' UTR, and, together with Hfq, inhibits the expression of two paralogous *fadE* mRNAs. The *fabB* and *fadE* genes are antagonistically regulated by the major fatty acid transcription factor, FadR, and we show that, together, FadR, FarS, and FadE constitute a mixed feed-forward loop regulating the transition between fatty acid biosynthesis and degradation in *V. cholerae*. Our results provide the molecular basis for studies on Hfq in *V. cholerae* and highlight the importance of a previously unrecognized sRNA for fatty acid metabolism in this major human pathogen.**

small RNA | feed-forward loop | fatty acid metabolism | RNase E | *Vibrio cholerae*

Many if not all microorganisms use posttranscriptional control mechanisms to regulate gene expression. Small regulatory RNAs (sRNAs) are frequently involved in these processes, and an overwhelming majority of sRNAs seem to function by base-pairing with either *cis*- or *trans*-encoded target transcripts. However, these sRNAs typically do not act in isolation but rather require the aid of RNA-binding proteins (1). One prime example of this type of proteins is the Hfq RNA chaperone. Hfq belongs to the family of Sm/Lsm proteins characterized by a multimeric, ring-like structure, which promotes the binding of nucleic acid molecules (2). Mechanistically, Hfq functions as a “molecular matchmaker” by facilitating the interaction of sRNAs with cognate target mRNAs. The protein also protects sRNAs from ribonucleolytic decay (3, 4). Hfq can make contact with RNA at four different sites—rim, distal face, proximal face, and C terminus—though not all Hfq homologs carry the C-terminal extension (5, 6).

Studies from bacterial model organisms such as *Escherichia coli* and *Salmonella enterica* showed that Hfq binds hundreds of mRNAs and several dozen sRNAs in vivo (7–10). Accordingly, deletions of *hfq* give rise to drastic phenotypic changes ranging from impaired stress responses to failure to engage collective cell functions, such as biofilm formation (11, 12). Significantly reduced infectivity is also observed for *hfq* mutants of pathogenic microorganisms (13), including the major human pathogen *Vibrio cholerae* (14). Here, activation of virulence gene expression relies on a complex pathway integrating signals from *V. cholerae* itself, other microorganisms, and the host (15, 16). Indeed, recent work on *V. cholerae*'s cholera toxin (CTX) has

revealed that host-derived heme and fatty acids are central factors for efficient colonization of the intestine (17).

Fatty acids also modulate the activity of the major virulence transcription factor ToxT (18), which, among many other genes, controls the expression of the TarB sRNA (19). TarB is a post-transcriptional inhibitor of the secreted colonization factor TcpF (19), as well as the pathogenicity island-encoded transcription factor VspR (20). In addition, *V. cholerae* sRNAs controlling cell-cell communication, e.g. Orr1-4 (21) and VqmR (22–24), as well as sRNAs responding to cell-envelope damage (25, 26), contribute to virulence gene expression.

Numerous other sRNAs exist in *V. cholerae*. In fact, transcriptomic approaches have reported hundreds of uncharacterized sRNAs, including a large group of sRNAs originating from the 3' end of mRNAs (23, 27). Similar observations have been made for other Gram-negative bacteria (28); however, it is often unclear if and how these sRNAs participate in gene regulation. Knowledge about the interaction of an sRNA with an RNA-binding protein can provide strong hypotheses regarding their regulatory functions. For example, Hfq- and ProQ-dependent sRNAs are likely to engage base-pairing with other transcripts, whereas CsrA-dependent sRNAs typically act by protein sequestration (1). For *V. cholerae*, we currently lack this information.

In this work, we have performed RIP-seq (RNA immunoprecipitation followed by high-throughput sequencing) analysis

## Significance

**Bacteria constantly transition between conditions of feast and famine. Colonization of the human intestine by *Vibrio cholerae* is associated with a surge in host-derived fatty acids, demanding rapid regulation of fatty acid metabolism. Here, we provide evidence for an RNA-based mechanism controlling the expression of central fatty acid metabolism genes in response to changing external fatty acid concentrations. We identified a small regulatory RNA, FarS, which is processed from the 3' UTR of the *fabB* fatty acid biosynthesis gene and inhibits the production of proteins required for fatty acid degradation. Tight control of fatty acid biosynthesis and degradation is vital for all bacteria, and, in *V. cholerae*, FarS plays an important role in balancing these processes.**

Author contributions: M.H., K.S.F., and K.P. designed research; M.H., K.S.F., and J.R. performed research; M.H., K.S.F., and K.P. analyzed data; and M.H. and K.P. wrote the paper. The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, <https://www.ncbi.nlm.nih.gov/geo> (accession no. GSE140516).

<sup>1</sup>To whom correspondence may be addressed. Email: kai.papenfort@uni-jena.de.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1920753117/-DCSupplemental>.

First published March 19, 2020.

of Hfq in *V. cholerae*. We discovered 603 mRNAs and 82 sRNAs interacting with Hfq. A total of 25 of these sRNAs were previously unknown, and 44 sRNAs mapped to the 3' end of a coding sequence. One highly abundant and 3'-encoded sRNA was FarS (for fatty acid regulated sRNA; as detailed later), which we studied in more detail. We show that FarS is expressed from the 3'UTR of the *fabB* gene producing  $\beta$ -ketoacyl-ACP synthase, a key enzyme for initiating fatty acid biosynthesis. As the *farS* gene does not have its own promoter, expression depends on the FadR transcription factor (activating *fabB*), as well as RNase E-mediated processing of the *fabB* mRNA. Mature FarS base-pairs with and inhibits the expression of two paralogous *fadE* mRNAs encoding acyl-CoA dehydrogenase, which is the rate-limiting enzyme in fatty acid  $\beta$ -oxidation (29). Transcription of *fadE* is repressed by FadR (30), and, together, FadR, *fabB*-FarS, and *fadE* constitute a previously unknown type 3 coherent feed-forward loop (FFL) regulating the transition between fatty acid biosynthesis and degradation in *V. cholerae*.

## Results

**RIP-Seq Analysis of Hfq in *V. Cholerae*.** To identify the RNA ligands of Hfq in *V. cholerae*, we added the 3XFLAG epitope to the C terminus of the chromosomal *hfq* locus (*vc0347*) and tested protein production at various stages of growth. Hfq was produced under all tested conditions (SI Appendix, Fig. S1A), allowing us to employ a RIP-seq approach (9) to determine the set of Hfq-bound transcripts of *V. cholerae* cells cultivated to low (OD<sub>600</sub> of 0.2) and high cell densities (OD<sub>600</sub> of 2.0). Western blot analysis of the coimmunoprecipitated samples revealed specific enrichment of the Hfq::3XFLAG protein when compared to the negative control lacking the FLAG epitope (Fig. 1A). Likewise, the Hfq-dependent sRNA Orr4 (21) was strongly enriched in the Hfq::3XFLAG samples, verifying our approach (Fig. 1B; note that Orr4 is most strongly expressed at low cell densities). To obtain the full set of Hfq binding partners from both cell densities, we next converted the copurified RNAs (using the Hfq::3XFLAG strain as well as the untagged controls) into cDNA, followed by deep sequencing (31). We obtained 7.9 to 39.1 million reads for the individual libraries, of which 96.7 to 98.0% mapped to the *V. cholerae* N16961 genome (32) (SI Appendix, Table S1). As expected, the majority of reads (~83%) obtained from the control libraries mapped to rRNAs, tRNAs, and housekeeping RNAs (tmRNA, 6S RNA, 4.5S RNA), whereas only 64% of reads mapped to this category in the Hfq::3XFLAG libraries (Fig. 1C). Instead, the fractions of bound mRNAs and sRNAs increased from 15 to 21% and from 2 to 15%, respectively. We detected a total of 82 sRNAs, 3 annotated riboswitches, and 603 mRNAs interacting with Hfq (>twofold enrichment over the untagged control sample; SI Appendix, Table S2). A total of 25 of these sRNAs (SI Appendix, Table S3) were discovered by our approach. As proof of concept, we confirmed that all previously reported Hfq-dependent sRNAs in *V. cholerae*, i.e., VqmR, Orr1-4, MicV, VrrA, MicX, RybB, TfoR, and TarA (23, 25, 33–37), were included in our dataset (SI Appendix, Table S2).

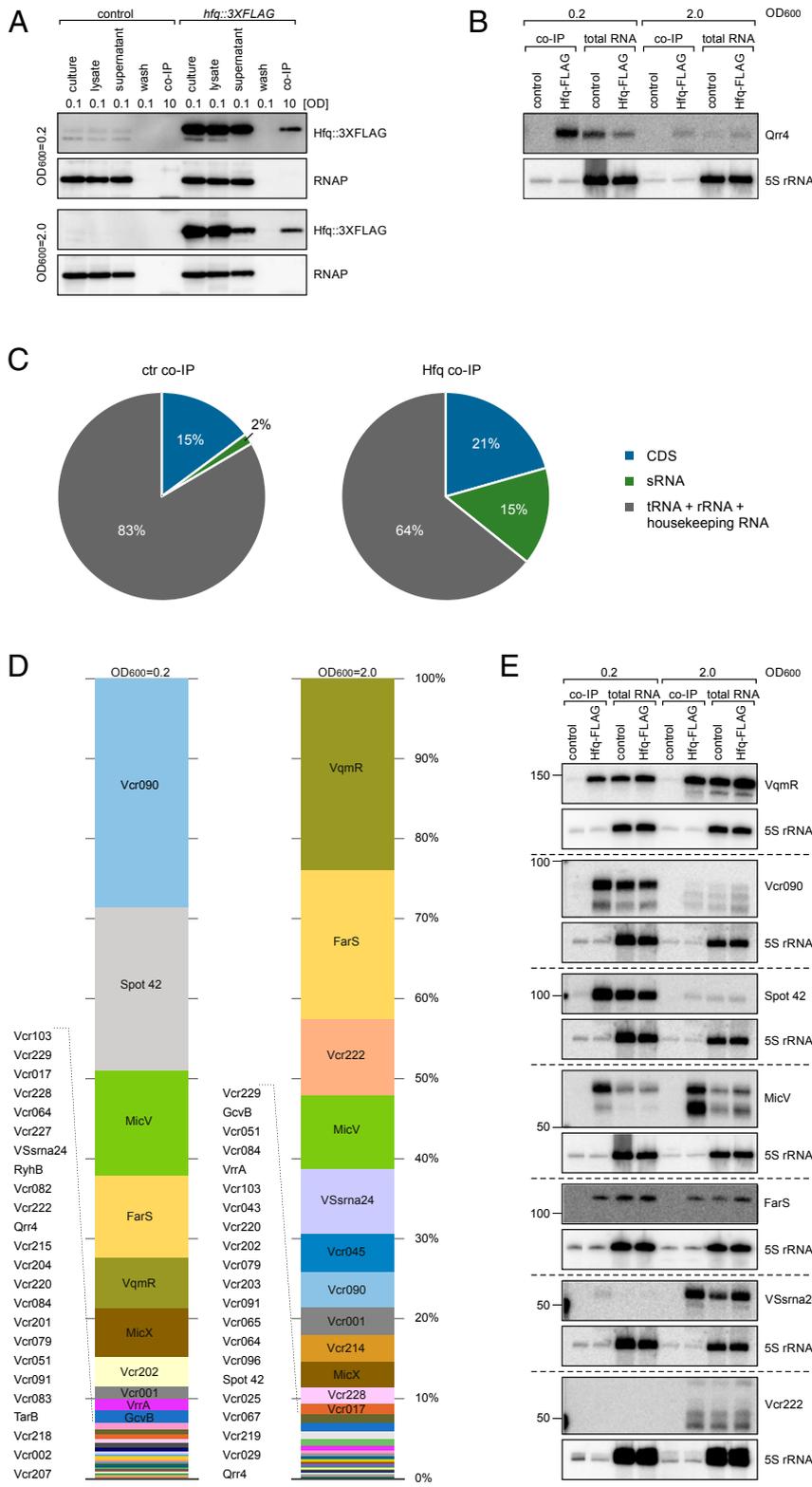
**Patterns of Hfq-Binding sRNAs at Low and High Cell Densities.** Next, we sorted the Hfq-binding sRNAs by abundance, i.e., the relative number of reads obtained from the Hfq::3XFLAG samples (Fig. 1D). At low cell densities, the top five most abundant sRNAs were the yet-uncharacterized Vcr090 sRNA (23), the highly conserved Spot 42 (38), MicV (26), FarS [previously identified as Vcr076 (23)], and VqmR (23). At high cell densities, the relative levels of Vcr090, MicV, and Spot 42 decreased, while VqmR became the most abundant sRNA, followed by FarS. The top five sRNAs now also included the carbon controlled VSSrna24 (39) and the newly discovered Vcr222 (Fig. 1D and SI Appendix, Table S3). We verified direct Hfq binding of these and 13 additional sRNAs using coimmunoprecipitation followed by Northern

blot analyses (Fig. 1E and SI Appendix, Fig. S1B). In addition, we performed electrophoretic mobility shift assays using purified Hfq and synthetic Vcr090, Spot 42, FarS, VqmR, MicV, and Vcr222 transcripts, which confirmed Hfq binding of these sRNAs in vitro (SI Appendix, Fig. S2 A–F). VqmR and Vcr090 displayed the highest affinity for Hfq in these assays ( $K_d$  of ~5 to 10 nM), while MicV showed the weakest binding ( $K_d$  of ~60 nM). These values are similar to Hfq-binding affinities of previously reported sRNAs, such as RybB and RydC from *S. enterica* (40, 41).

**An Abundant Class of 3'UTR-Derived Hfq-Binding sRNAs.** Our previous transcriptome analysis of *V. cholerae* cultivated under conditions of low and high cell densities indicated 44 possible 3'UTR-derived sRNAs (23); however, it remained unclear if these sRNAs were involved in posttranscriptional gene control and if Hfq would be required in this process. To address this question, we categorized the list of Hfq-binding sRNAs by their genomic location, i.e., intergenic, 5'UTR, CDS, and 3'UTR (Fig. 2A). In line with our previous hypothesis, we discovered that a large fraction of Hfq-binding sRNAs are expressed from the 3'UTR of mRNAs (54%), followed by intergenic sRNAs (37%) and sRNAs located in 5'UTRs (8%). Only one sRNA originated from an annotated coding sequence.

These results suggested that 3'UTR-derived sRNAs could have important regulatory roles in *V. cholerae*. To test this prediction, we focused on FarS, which was the most abundant 3'UTR-derived sRNA in our Hfq coimmunoprecipitation experiments (Fig. 1D). The *farS* gene is located in the 3'UTR of *fabB* (encoding  $\beta$ -ketoacyl-ACP synthase; Fig. 2B) and highly conserved among the *Vibrio* spp. (Fig. 2C). Northern analysis of *V. cholerae* cultivated in rich medium indicated that FarS is detectable at all stages of growth (Fig. 2D, lanes 1 to 4), and similar results were obtained for growth in minimal medium (SI Appendix, Fig. S3A, lanes 1 to 4). Expression of *fabB* has previously been reported to rely on the dual transcriptional regulator FadR (42), and we were able to confirm this result (SI Appendix, Fig. S3B). We therefore speculated that FadR might also affect the expression of FarS. Indeed, mutation of *fadR* resulted in approximately eightfold reduced *farS* levels in rich and minimal medium (Fig. 2D and SI Appendix, Fig. S3A, lanes 5 to 8), and expression was fully complemented by introduction of an FadR-producing plasmid (Fig. 2D and SI Appendix, Fig. S3A, lanes 9 to 12). These results are in agreement with a previous study suggesting that *farS* does not have its own promoter (23) and indicated that FarS expression strictly relies on transcriptional input signals integrated at the *fabB* promoter. To test this hypothesis, we first constructed a *farS* mutant strain by removing base pairs 1 to 85 of the *farS* sequence from the *V. cholerae* genome while keeping the Rho-independent terminator intact (Fig. 2C). Importantly, this mutation did not affect *fabB* mRNA stability (SI Appendix, Fig. S3C), avoiding possible secondary effects resulting from this mutation. We next introduced a plasmid containing the *fabB*-*farS* gene locus, as well as the *fabB* promoter, and monitored FarS production by Northern analysis. As expected, the *fabB*-*farS* plasmid fully restored FarS expression in the  $\Delta farS$  mutant (Fig. 2E, lanes 1 to 3). In contrast, deletion of the *fabB* promoter sequence in the *fabB*-*farS* plasmid strongly reduced FarS levels (>100-fold; Fig. 2E, lane 4), and expression remained low when we eliminated additional segments of the *fabB* coding sequence in the *fabB*-*farS* plasmid (Fig. 2E, lanes 5 and 6). Together, these results show that FarS is produced from the 3'UTR of *fabB* and that expression of the sRNA depends on the *fabB* promoter.

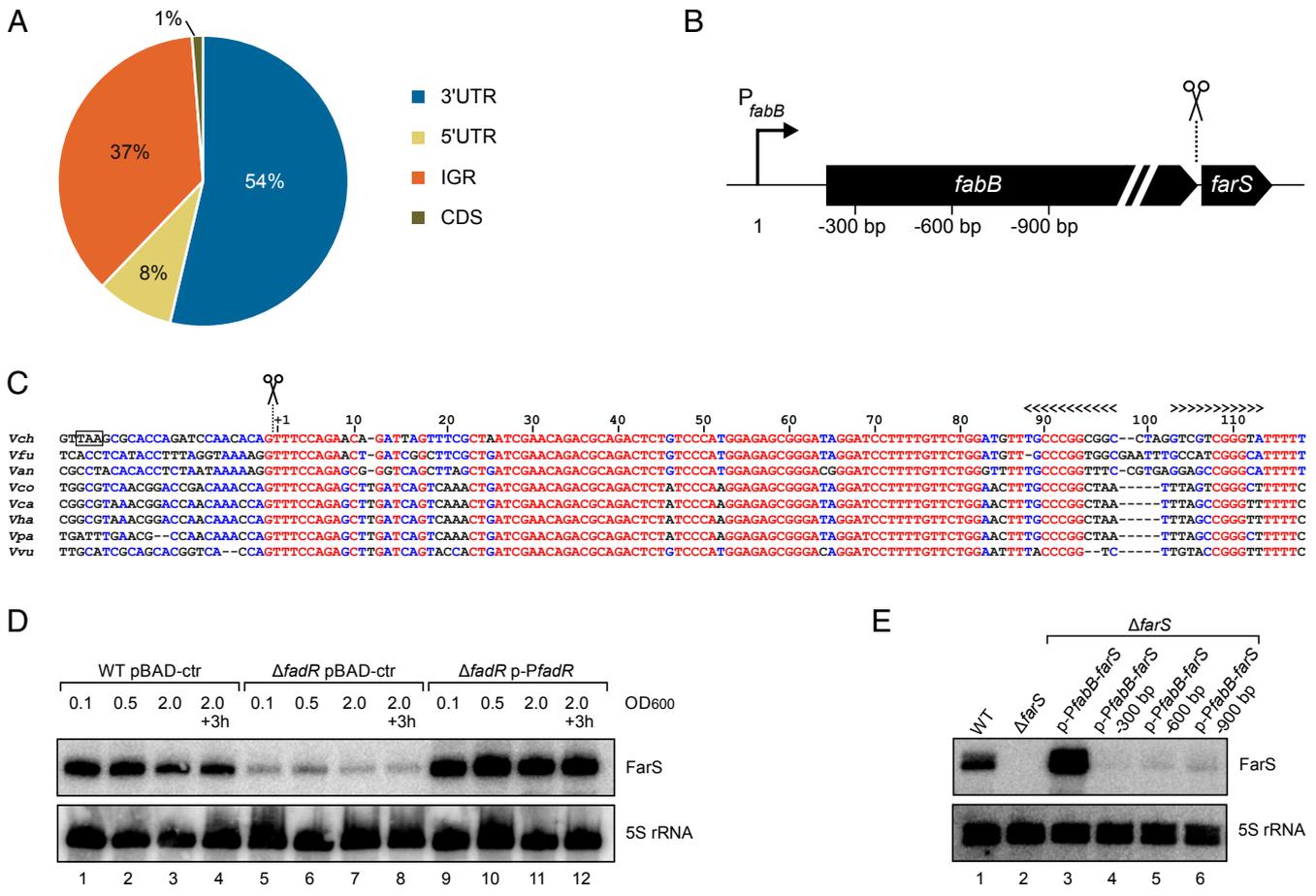
**RNase E Is Required for FarS Production.** The class of 3'UTR-derived sRNAs has been divided into two groups: (i) sRNAs carrying their promoters [e.g., DapZ (9) and MicL (43)] and (ii) sRNAs requiring ribonuclease-dependent cleavage for full maturation [e.g., SdhX (44, 45) and CpxQ (46, 47)]. Our previous results



**Fig. 1.** RIP-seq analysis of Hfq-binding sRNAs. (A) *V. cholerae* wild-type cells (control) and cells carrying a 3XFLAG epitope at the C-terminal end of the chromosomal *hfq* gene were cultivated in LB medium to low (OD<sub>600</sub> of 0.2) and high cell densities (OD<sub>600</sub> of 2.0) and subjected to coimmunoprecipitation. Protein samples were collected at different steps of the IP procedure and analyzed by Western blots. Culture refers to total protein before treatment, lysate refers to total protein after cell lysis, supernatant refers to remaining protein after incubation with anti-FLAG antibody and protein G Sepharose, wash refers to remaining protein in the lysis buffer after five washing steps, and co-IP indicates coimmunoprecipitated protein sample. The relative amount of cells loaded (OD<sub>600</sub> units) is indicated. RNAP served as loading control. (B) RNA samples of co-IP and total RNA (lysate) fractions were loaded on a Northern blot and analyzed for Qrr4 levels. 5S rRNA served as loading control. (C) Pie charts of control and *Hfq* co-IP samples showing the relative fractions of the different RNA classes. The relative amount of total cDNA reads from each class in the control and *Hfq* co-IP libraries are shown. (D) Distribution of reads of significantly enriched sRNAs (fold enrichment > 2, *P* value ≤ 0.05) in *Hfq* co-IP libraries obtained from low (OD<sub>600</sub> of 0.2) and high cell densities (OD<sub>600</sub> of 2.0). Reads matching to a given sRNA were compared to all enriched sRNAs in the cDNA libraries. Shown are all sRNAs corresponding to at least 0.1% of the mapped reads. The relative amount of reads and enrichment factors for each sRNA are listed in [SI Appendix, Table S2](#). (E) Co-IP and total RNA (lysate) fractions were obtained from *V. cholerae* wild-type and *hfq::3XFLAG*-tagged strains cultivated in LB medium to low (OD<sub>600</sub> of 0.2) and high cell densities (OD<sub>600</sub> of 2.0). The RNA was loaded on Northern blots and probed for the indicated sRNAs. 5S rRNA served as a loading control.

indicated that FarS belongs to the second class (Fig. 2 D and E); however, the respective ribonuclease required for FarS maturation remained unclear. Inspection of the *farS* gene revealed a conserved sequence stretch located at the very 5' end of the sRNA (Fig. 2C), matching the recently determined recognition motif for RNase E-mediated cleavage (48). To test a possible involvement of RNase E in FarS maturation, we transferred the *farS* mutation into a

*V. cholerae* strain, producing a temperature-sensitive RNase E variant [*rne* encodes RNase E and is an essential gene in *V. cholerae* (49)], and transformed this strain with a plasmid allowing pBAD-inducible expression of the *fabB-farS* gene locus. We cultivated this strain under permissive (30 °C) and nonpermissive temperature (44 °C) and induced the pBAD promoter by addition of L-arabinose (0.2% final concentration). Total RNA samples



**Fig. 2.** Identification and expression of the FarS sRNA. (A) Classification of Hfq-binding sRNAs according to their genomic location. The pie chart shows the relative fractions of Hfq-binding sRNAs (fold enrichment > 2, *P* value ≤ 0.05) originating from 3' UTRs, intergenic regions (IGRs), 5' UTRs, and coding sequences (CDSs). (B) Schematic representation of the *fabB-farS* genomic organization. Scissors indicate the processing site. Numbers correspond to the *fabB* promoter truncations tested in *E. coli*. (C) Alignment of *farS* sequences in different *Vibrio* species. The sequences were aligned using the Multalign tool (76). The start of the sRNA and the Rho-independent terminator are indicated. The stop codon of *fabB* in *V. cholerae* is marked with a black box. *Vch*, *Vibrio cholerae*; *Vfu*, *Vibrio furnissii*; *Van*, *Vibrio anguillarum*; *Vco*, *Vibrio coralliilyticus*; *Vca*, *Vibrio campbellii*; *Vha*, *Vibrio harveyi*; *Vpa*, *Vibrio parahaemolyticus*; *Vvu*, *Vibrio vulnificus*. (D) *V. cholerae* wild-type and  $\Delta$ *fadR* cells harboring either a control plasmid (pBAD-ctr) or a plasmid containing the *fadR* gene and its native promoter (p-P*fadR*) were cultivated in LB medium. Total RNA samples were collected at different stages of growth, and expression of FarS was analyzed on Northern blot. 5S rRNA was used as loading control. (E) *V. cholerae* wild-type and  $\Delta$ *farS* strains harboring different plasmids containing *fabB-farS* gene fragments (as indicated in B) were grown to stationary phase (OD<sub>600</sub> of 2.0) in LB medium. Northern blot analysis was performed to determine FarS levels. Probing for 5S rRNA served as a loading control.

were collected, and FarS expression was tested by Northern analysis. Mature FarS expression was readily detected at 30 °C but strongly reduced at 44 °C in the temperature-sensitive RNase E mutant (Fig. 3A, lanes 3 and 4). In addition, nonpermissive temperatures also resulted in the accumulation of various processing intermediates, suggesting inadequate degradation (maturation) of the *fabB* mRNA. This effect was specific to RNase E, as the relevant control strain (carrying the native *me* gene) displayed accurate FarS maturation at 30 °C and 44 °C (Fig. 3A, lanes 1 and 2).

To corroborate a direct role of RNase E in FarS production, we exchanged the first three base pairs of *farS* (TTT to GGG; mutating the predicted RNase E recognition motif) and tested FarS production. In line with our prediction, exchange of these critical residues almost completely abolished FarS production, while cleavage events located further upstream in the transcript remained functional (Fig. 3B). In summary, these data strongly suggest that FarS is produced by RNase E-mediated processing of the *fabB* mRNA.

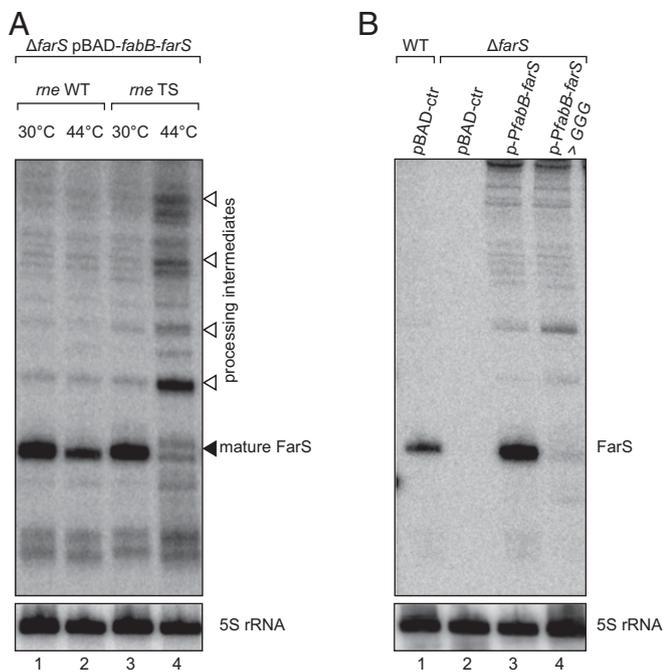
**FarS inhibits the Expression of Two Paralogous *fadE* mRNAs. A hallmark of Hfq-dependent sRNAs is their ability to base-pair with *trans*-encoded target mRNAs, affecting transcript stability**

and translation initiation (3). This feature has been demonstrated for conventional sRNAs encoded by free-standing genes, as well as 3'UTR-derived sRNAs requiring ribonuclease-assisted maturation (28). To investigate if FarS functions as a *trans*-acting regulator in *V. cholerae*, we cloned the *farS* gene onto a plasmid downstream of the pBAD promoter (initiating transcription at the RNase E cleavage site; see Fig. 2C). Next, we cultivated *V. cholerae* wild-type cells carrying either pBAD-*farS* or a control plasmid to exponential phase (OD<sub>600</sub> of 0.5) and induced pBAD expression

**Table 1. Genes differentially expressed in response to FarS pulse expression**

Gene	Description	Fold change
<i>vc1740</i>	acyl-CoA dehydrogenase	-2.02
<i>vc2231</i>	acyl-CoA dehydrogenase	-2.54

Description is based on the annotation at KEGG (<https://www.genome.jp/kegg/>). Fold changes were obtained by transcriptomic analysis of pBAD-driven FarS expression using RNA-seq. Genes regulated >twofold with an FDR-adjusted *P* value ≤ 0.05 are listed.

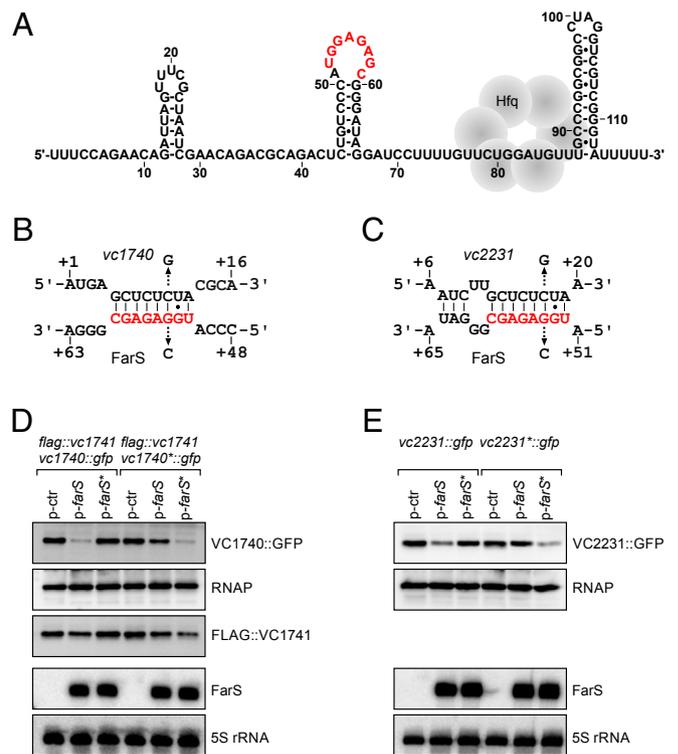


**Fig. 3.** RNase E is required for FarS production. (A) *V. cholerae*  $\Delta farS$  carrying either a wild-type (*rne* WT) or temperature-sensitive RNase E (*rne* TS) allele and the pBAD-*fabB-farS* plasmid were cultivated at 30 °C in LB medium. When cells reached an OD<sub>600</sub> of 1.0, cultures were split and kept at permissive temperature (30 °C) or shifted to nonpermissive temperature (44 °C) and incubated for 30 min. Next, expression of *fabB-farS* was induced using L-arabinose (0.2% final concentration, 30 min), and FarS levels were monitored by Northern blot. Probing for 5S rRNA served as loading control. (B) *V. cholerae* wild-type or  $\Delta farS$  cells harboring either a control plasmid (pBAD-ctr), a plasmid containing the *fabB-farS* gene locus and the *fabB* promoter, or a version of the plasmid where the first three base pairs of *farS* were mutated (TTT to GGG) were cultivated in LB medium. Total RNA samples were collected when cells reached an OD<sub>600</sub> of 1.0. The Northern blot was probed for FarS, and 5S rRNA was used as a loading control.

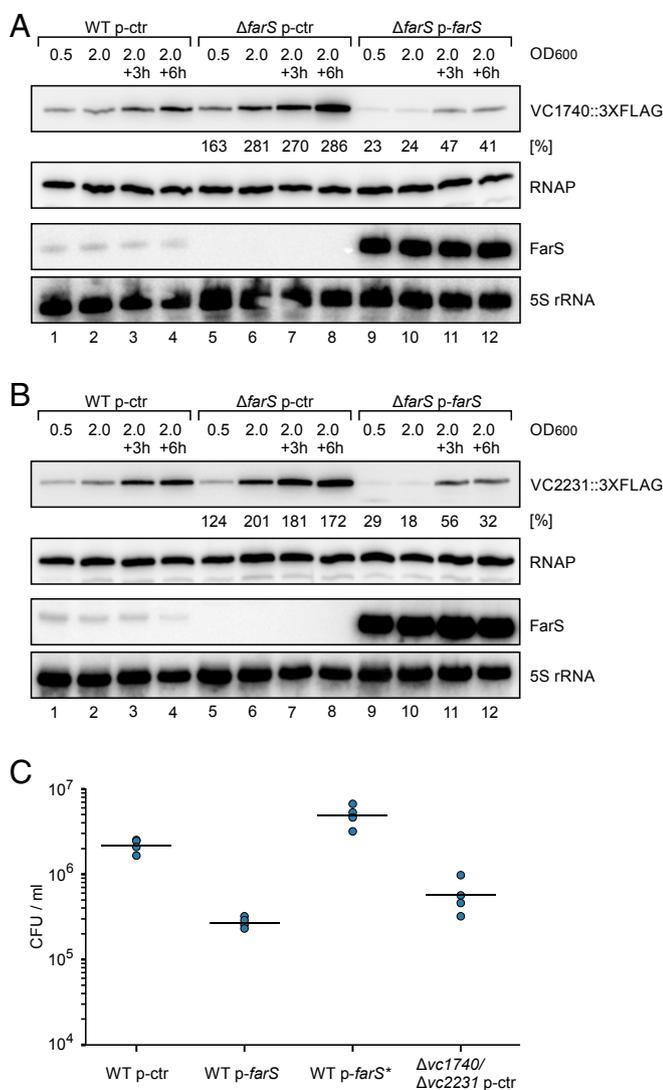
for 15 min. Comparison of pBAD-controlled and endogenous FarS expression revealed that both transcripts migrated at the same size and that pBAD-driven FarS levels were ~10-fold increased when compared to the control (SI Appendix, Fig. S4A). Next, we converted the total RNA obtained from these samples into cDNA and determined global transcriptome changes using deep sequencing (31). We identified only two differentially expressed transcripts upon FarS induction (both repressed, *vc1740* and *vc2231*; Table 1), and both encoded homologs of the same enzyme, FadE (acyl-CoA dehydrogenase).

To explore the mechanistic details of FarS-mediated *fadE* repression, we first determined the secondary structure of FarS using chemical and enzymatic probing (SI Appendix, Fig. S4B). FarS contains three hairpin elements, the most distal of which most likely serves as a Rho-independent terminator (Fig. 4A). In addition, we identified a potential Hfq-binding site located between the second and third stem-loops of FarS, indicating that base-pairing will occur further upstream in the sRNA. Using the RNAhybrid algorithm (50), we were able to predict potential base-pairing sites of FarS with *vc1740* and *vc2231* involving an exposed loop in the second hairpin element of FarS (Fig. 4A) and the sequence encoding the N terminus of both FadE proteins (Fig. 4B and C). To test these predictions, we used *E. coli* as a heterologous host (lacking the *farS* gene) and a reporter system with constitutive expression of the sRNA and a translational fusion of the target mRNA to *gfp* on two individual plasmids (51). Given that *vc1740* constitutes the second gene in a di-cistronic operon with

*vc1741* (23) (encoding a transcriptional regulator), we employed a variant of this system in which the proximal gene (i.e., *vc1741*) carries an N-terminal FLAG epitope to monitor the effect of the sRNA on both genes. Using this setup, we discovered that FarS specifically repressed the distal part of this operon (producing FadE::GFP), whereas FLAG::VC1741 levels remained unchanged (Fig. 4D). We also found that FarS inhibits VC2231::GFP expression at the posttranscriptional level (Fig. 4E) and that a single point mutation in FarS (G54 to C) was sufficient to block regulation of both *fadE* targets (Fig. 4D and E). Vice versa, mutation of *vc1740* and *vc2231* at the indicated positions (C10 to G and C17 to G, respectively; Fig. 4B and C) inhibited regulation by FarS, while combination of the mutated FarS and *fadE* variants restored GFP repression (Fig. 4D and E). FarS stability was reduced in *hfq*-deficient *V. cholerae* cells (SI Appendix, Fig. S5A), and, despite FarS accumulating to similar levels when expressed from a multicopy plasmid in wild-type and  $\Delta hfq$  cells, regulation of both targets was strictly dependent on Hfq (SI Appendix, Fig. S5B and C). Our data suggest that FarS uses an exposed loop element to base-pair with the two paralogous *fadE* mRNAs and that this process requires Hfq. We note that, although



**Fig. 4.** Structure of FarS and base-pairing to *fadE* target mRNAs. (A) Secondary structure of FarS. The secondary structure was derived from chemical and enzymatic structure probing experiments (SI Appendix, Fig. S4B). The base-pairing site is marked in red, and the Hfq binding site is indicated. (B) Predicted base-pairing of FarS with *vc1740*. Arrows indicate the single nucleotide mutations tested in D. (C) Predicted base-pairing of FarS with *vc2231*. Arrows indicate the single nucleotide mutations tested in E. (D) Discoordinate regulation of FLAG::VC1741 and VC1740::GFP. *E. coli* cells carrying a reporter plasmid for FLAG::VC1741 and VC1740::GFP or VC1740\*::GFP (C10 to G) were cotransformed with a control plasmid (p-ctr), p-*farS*, or p-*farS*\* (G54 to C). Cells were grown in LB medium to stationary phase (OD<sub>600</sub> of 2.0). GFP and FLAG levels were measured by Western blot, and FarS levels were determined by Northern blot. RNAP and 5S rRNA served as loading controls for the Western and Northern blots, respectively. (E) *E. coli* harboring a reporter plasmid for VC2231::GFP or VC2231\*::GFP (C17 to G) and either a control plasmid (p-ctr), p-*farS*, or p-*farS*\* (G54 to C) were grown in LB medium to OD<sub>600</sub> of 2.0. GFP and FarS levels were monitored as in D.



**Fig. 5. FarS inhibits FadE protein production.** (A and B) *V. cholerae* wild-type and  $\Delta farS$  strains carrying a chromosomal 3XFLAG epitope either at the *vc1740* (A) or at the *vc2231* (B) gene and harboring the indicated plasmids were cultivated in M9 minimal medium. Protein and total RNA samples were collected at the indicated OD<sub>600</sub> readings. FadE::3XFLAG protein production (A, VC1740::3XFLAG; B, VC2231::3XFLAG) was analyzed on Western blots, and expression of FarS was monitored on Northern blots. RNAP and 5S rRNA served as loading controls for the Western and Northern blots, respectively. Percentages indicate the amount of protein relative to the wild-type level at the corresponding growth phase. A quantification of data obtained from three independent biological replicates is shown in *SI Appendix, Fig. S6 A and B*. (C) *V. cholerae* wild-type and  $\Delta vc1740/\Delta vc2231$  strains carrying the indicated plasmids were cultivated for 10 h in M9 minimal medium containing fatty acid (sodium oleate) as sole carbon source. Serial dilutions were prepared and recovered on agar plates, and colony-forming units (CFU) per milliliter were determined. Dots represent individual replicates ( $n = 4$ ), and lines indicate the mean CFU.

the FadE protein sequence is highly divergent at the N terminus, the FarS base-pairing site is highly conserved at the DNA level (*SI Appendix, Fig. S5 D and E*). These results might suggest that, at the phylogenetic level, FarS-mediated repression of the *fadE* mRNA was established before the gene was duplicated.

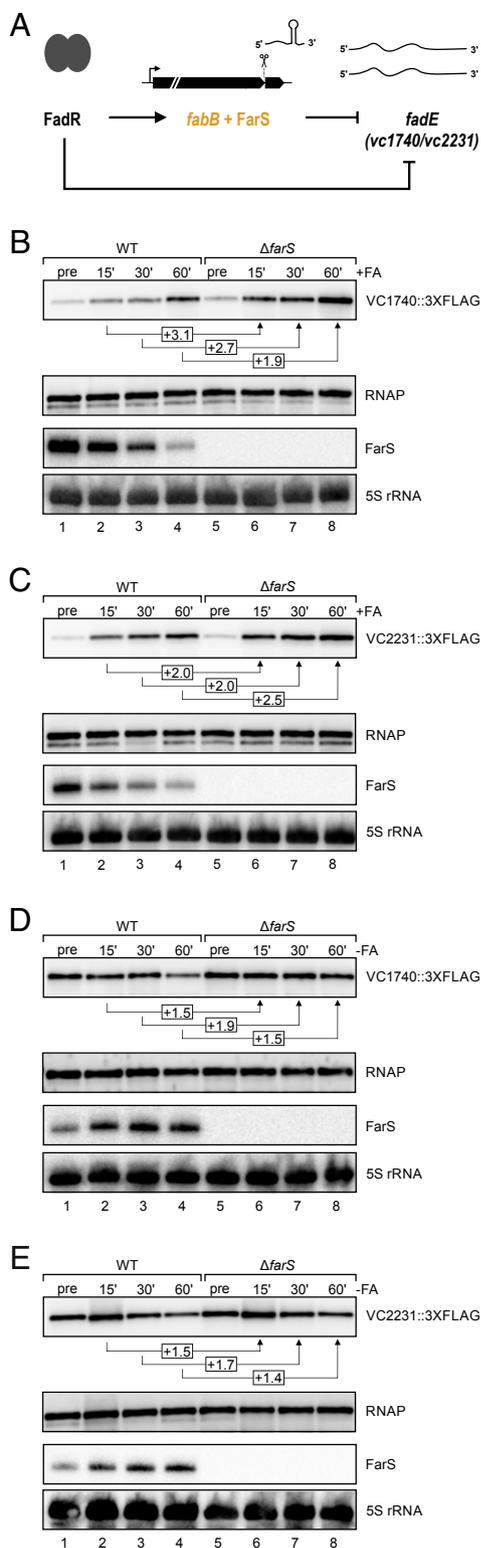
**FarS Restricts FadE Protein Production.** Repression of the two *fadE::gfp* fusions suggested that FarS also inhibits the synthesis of both FadE paralogs in *V. cholerae*. To test this hypothesis, we

added a 3XFLAG epitope to the C termini of the chromosomal *vc1740* and *vc2231* genes and monitored protein production in wild-type and  $\Delta farS$  cells (both harboring a control plasmid). In agreement with our prediction, the production of both proteins was elevated in *farS*-deficient cells. For VC1740::3XFLAG, increased protein abundance was detected at all stages of growth accumulating to ~2.5-fold higher levels in late stationary phase (6 h after cells reached an OD<sub>600</sub> of 2.0; Fig. 5A, lanes 1 to 8, and *SI Appendix, Fig. S6A*). Similarly, VC2231::3XFLAG levels were elevated in  $\Delta farS$ , with the most pronounced differences in protein production (~twofold) when cells reached an OD<sub>600</sub> of 2.0 (Fig. 5B, lanes 1 to 8, and *SI Appendix, Fig. S6B*). In both cases (VC1740::3XFLAG and VC2231::3XFLAG), introduction of a FarS overexpression plasmid into  $\Delta farS$  cells strongly reduced FadE levels at all stages of growth (Fig. 5A and B, lanes 9 to 12, and *SI Appendix, Fig. S6 A and B*).

To test the effect of FarS-mediated FadE repression on fatty acid metabolism of *V. cholerae*, we cultivated wild-type *V. cholerae* carrying either a control or the FarS overexpression plasmid in minimal medium containing sodium oleate as sole carbon source. We discovered that, after 10 h of incubation under these conditions, *V. cholerae* cells expressing FarS from a plasmid displayed ~10-fold decreased survival when counted on agar plates (Fig. 5C). This effect was specific to the repression of *vc1740* and *vc2231* by FarS, since plasmid-borne expression of mutated FarS (FarS\*, e.g., see Fig. 4B) did not inhibit growth under these conditions, whereas *V. cholerae* cells deleted for *vc1740* and *vc2231* showed survival rates similar to the FarS overexpression strain. Together, we conclude that FarS down-regulates the synthesis of both FadE paralogs and thereby affects the fatty acid metabolism in *V. cholerae*.

**FarS Is the Central Regulator of a Mixed Feed-Forward Loop.** Previous reports have shown that FadR of *V. cholerae* functions as a dual transcriptional regulator inhibiting *fadE* and activating *fabB* (52). Our data now show that FarS is coexpressed from the *fabB* promoter (Fig. 2E) and represses the production of the FadE paralogs (Fig. 5A and B and *SI Appendix, Fig. S6 A and B*). Thus, FadR and FarS both repress synthesis of the FadE proteins, establishing a mixed type 3 coherent FFL in which one regulator, FarS, is clipped off the 3'UTR of a functionally related mRNA product (Fig. 6A). The logic implied in this regulatory setup predicts two possible functions for FarS: (i) FarS acts as a delay element limiting FadE production when *V. cholerae* transitions from low to high external fatty acid concentrations and, (ii) in the reverse scenario (transition from high to low fatty acid concentrations), FarS accelerates the repression of FadE. To test this prediction, we examined the effects of adding or removing fatty acids. We first determined the effects on the sRNA levels by cultivating *V. cholerae* cells to early stationary phase (OD<sub>600</sub> of 1.0) and then monitoring the expression of FarS in response to the addition or removal of fatty acids. Indeed, addition of sodium oleate (0.005% final concentration) efficiently repressed FarS production in *V. cholerae* (~fivefold; *SI Appendix, Fig. S6C*). Removal of external fatty acids from the medium, on the contrary (by washing and reinoculation of *V. cholerae* cells into fatty-acid-free minimal medium), resulted in increased FarS expression (~10-fold 60 min after reinoculation; *SI Appendix, Fig. S6D*). These results are in line with the expected patterns of FabB and FarS expression under conditions of low and high fatty acids.

We next tested the effect of fatty acid addition and removal on the expression of the two FadE paralogs in wild-type and  $\Delta farS$  *V. cholerae* using the experimental conditions established earlier. We discovered that addition of sodium oleate led to the increased production of the VC1740 and VC2231 proteins, albeit with slightly different kinetics. Accumulation of VC2231 was more rapid when compared to VC1740, and VC2231 showed a larger dynamic range (~12-fold vs. ~8-fold increased protein production comparing the



**Fig. 6.** FarS is part of a mixed feed-forward loop. (A) Schematic display of a mixed type 3 coherent feed-forward loop involving the transcription factor FadR, the *fabB* mRNA, FarS, and the two *fadE* mRNAs. (B and C) *V. cholerae* wild-type and  $\Delta farS$  strains carrying a chromosomal 3XFLAG epitope either at the *vc1740* (B) or at the *vc2231* (C) gene were cultivated in M9 minimal medium to stationary phase (OD<sub>600</sub> of 2.0). Total protein and RNA samples were collected before and after addition of fatty acids (+FA; sodium oleate, 0.005% final concentration) at the indicated time points. Expression patterns of the VC1740 (B) and VC2231 (C) proteins were analyzed on Western blots, and expression of FarS was determined

preinduction and the 60-min time points; Fig. 6 B and C, lanes 1 to 4). *V. cholerae* cells lacking *farS* also showed elevated VC1740 and VC2231 levels upon fatty acid supplementation; however, the dynamics of the response were accelerated and resulted in ~2 to 2.5-fold higher protein levels at the final time point of the experiment (Fig. 6 B and C, lanes 5 to 8, and *SI Appendix, Fig. S6 E and F*). Thus, we conclude that, as hypothesized earlier, FarS slows down FadE protein production when *V. cholerae* is exposed to sudden surges in fatty acid concentration.

FarS also inhibited FadE production when fatty acids were removed from the environment. Here, we observed that VC1740 and VC2231 levels decreased upon reinoculation of *V. cholerae* into fresh medium lacking fatty acids and that  $\Delta farS$  cells displayed ~1.5 to 2-fold higher protein levels during the course of the experiment (Fig. 6 D and E and *SI Appendix, Fig. S6 G and H*). Of note, FarS only acts to down-regulate the existing *vc1740* and *vc2231* mRNAs, while transcription of these genes is simultaneously repressed by FadR (52). Together, both factors (transcriptional and posttranscriptional regulation) allow FadE repression when fatty acids become scarce.

## Discussion

Bacterial sRNAs constitute a heterogeneous group of regulators that are produced from almost all segments of the genome. Traditionally, sRNAs from intergenic regions have been the focus of attention, which may well be explained by the design of early biocomputational screens scoring for conserved sequences associated with potential promoters and Rho-independent terminators that should be transcribed independent of both adjacent genes (53). Similarly, microarray-based approaches discovered a wealth of sRNAs in model organisms such as *E. coli* (8); however, due to the lacking resolution in microarray technologies, these analyses also favored the discovery of sRNAs from intergenic regions. The perception that sRNAs strictly originate from intergenic sequences was first challenged by shotgun cloning approaches (54) and further revised using deep-sequencing analyses (9, 48, 55).

In the well-studied *Salmonella* and *E. coli* models, 3'UTR-derived sRNAs constitute ~20 to 30% of the Hfq-binding sRNAs, and, in this study, we reveal even higher numbers for *V. cholerae* (54%; Fig. 2A). The molecular determinants for this strong preference for 3'UTRs might well be explained by Hfq's affinity toward Rho-independent terminators (56) and its relatively weak sequence specificity (57). However, the TransTerm algorithm (58) predicts a total of ~760 high-confidence Rho-independent terminators for *V. cholerae*, suggesting that additional factors are required to guide Hfq to these 3'UTRs. It is interesting to note that initial FarS biogenesis seems to be independent of Hfq, as mature FarS is readily detectable in  $\Delta hfq$  cells (*SI Appendix, Fig. S5A*). This finding could also indicate an order of events for the synthesis of 3'UTR-derived sRNAs in which transcription is followed by RNase E-mediated mRNA decay, which is followed by binding of Hfq to the final degradation product. This process differs from the reported maturation of the ArcZ sRNA. Here, Hfq binding to the sRNA's 3' end is

using Northern blot analysis. RNAP and 5S rRNA served as loading controls for the Western and Northern blots, respectively. (D and E) *V. cholerae* wild-type and  $\Delta farS$  strains carrying a chromosomal 3XFLAG epitope either at the *vc1740* (D) or at the *vc2231* (E) gene were cultivated in M9 minimal medium containing sodium oleate (0.005% final concentration) to an OD<sub>600</sub> of 2.0. Cells were washed with PBS and resuspended in M9 minimal medium lacking fatty acids (-FA). Total protein and RNA samples were collected before and after removal of fatty acids at the indicated time points. Western and Northern blots show VC1740 (D) and VC2231 (E) protein and FarS levels, respectively. RNAP was used as loading control for Western blots; 5S rRNA for Northern blots.

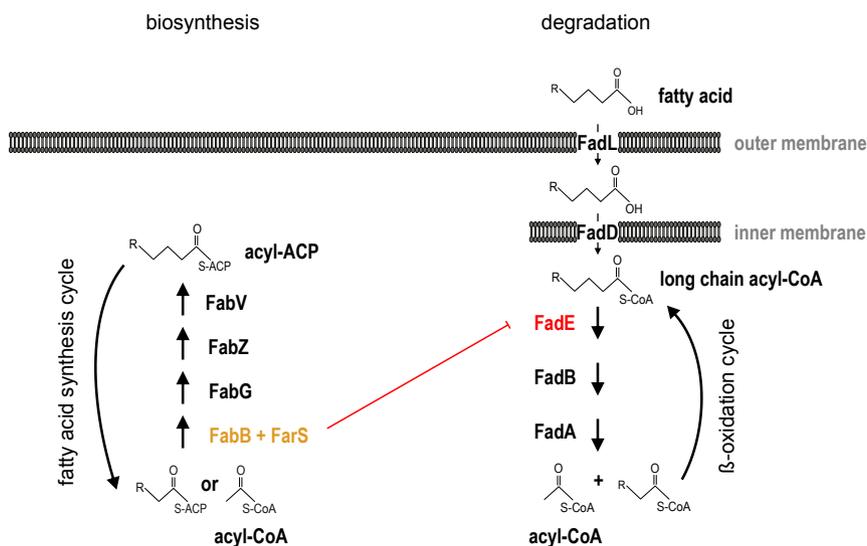
required to guide RNase E to the correct cleavage position (48). For 3'UTR-derived sRNAs such as MicL, CpxQ, and FarS, this function might be compensated by the presence of stem-loop elements inhibiting continued RNase E-mediated transcript decay (Fig. 4A, *SI Appendix*, Fig. S4B, and refs. 46 and 59).

Understanding the biogenesis and functions of 3'UTR-derived sRNAs could also make an interesting case for the study of sRNA evolution in bacteria (60). This is particularly evident for sRNAs that are produced by ribonucleolytic cleavage from mRNAs, as transcriptional control is already established by the promoter elements located upstream of coding sequence(s). Given that RNase E-mediated cleavage is pervasive in the enterobacteria (48, 61), one may speculate that de novo sRNA generation from 3'UTRs is driven by the affinity of Hfq for Rho-independent terminator elements (56) followed by potential base pair mutations that allow for the interaction with selected *trans*-encoded mRNAs. Other global RNA chaperones such as ProQ (55, 62, 63) could take Hfq's position in this scenario as well and mediate target mRNA interactions. For example, RaiZ of *S. enterica*, still the single thoroughly characterized ProQ-dependent sRNA, is produced by cleavage of the *raiA* mRNA and base-pairs with the *hupA* mRNA to repress translation initiation (64). Interestingly, while FarS produced from the *fabB* 3'UTR of *V. cholerae* binds Hfq (Figs. 1 and 2), the *fabB* 3'UTR is a strong binding partner of ProQ in *E. coli* and *Salmonella* (65), indicating two possible analogous pathways (using Hfq or ProQ) to evolve functional sRNAs from the 3' end of mRNAs. Finally, in contrast to their upstream coding sequences, the 3' ends of mRNAs typically do not show conservation at the sequence level unless these are required to base-pair with mRNAs (Fig. 2C and refs. 9, 34, 43–45, and 48). This might be an exploitable feature for future bioinformatic searches aiming at 3'UTR-derived sRNAs in other microbes that have not yet been investigated for their Hfq–RNA interactions.

The expression of a regulatory RNA from the 3' end of an mRNA using ribonucleolytic cleavage also adds an intriguing feature to the operon concept (54). Operons typically constitute a set of coding genes that are cotranscribed and together build a

biologically relevant unit or pathway. This concept has now been extended to noncoding regulators, as these can provide a regulatory function to mRNAs, which would typically only produce an enzyme or a structural protein. How the regulatory role of these 3' end-encoded sRNAs relates to the function of their upstream coding sequences has now been established in several cases. First, CpxQ is part of the CpxAR stress response system in *Salmonella* and produced from the 3'UTR of *cpxP* to reduce the translation of inner membrane proteins that trigger the same pathway (46). Second, SdhX, which is cotranscribed with the ~10-kb-long *sdhCDAB-sucABCD* operon of the TCA cycle, down-regulates the synthesis of AckA (acetate kinase) and thereby adjusts TCA flux and acetate metabolism (44, 45). Third, s-SodF sRNA is expressed from the 3'UTR of the *sodF* mRNA (encoding a Fe-containing superoxide dismutase) under nickel starvation and limits the synthesis of the nickel-containing SodN superoxide dismutase (66). We showed here that FarS (produced from the 3' end of the *fabB* fatty acid biosynthesis gene) inhibits the expression of two paralogous FadE proteins, which are involved in fatty acid degradation (Fig. 7). Thus, in the studied examples, the 3' end-derived sRNAs provide a strong functional link between their origin of expression (i.e., their upstream mRNAs) and their targets. Importantly, s-SodF is produced in Gram-positive *Streptomyces coelicolor*, and regulation of *sodN* does not require Hfq (66), suggesting that this type of gene control is relevant beyond the enterobacterial clade. Stable 3'UTR RNA tails have now also been documented in mammalian cells (67), proposing an even broader regulatory concept.

One exciting future question related to the biological roles of 3'UTR-derived sRNAs is how they modulate the dynamics of their associated regulatory systems or pathways. We could show here that FarS is part of a type 3 coherent FFL that modifies the expression of two *fadE* genes (i.e., *vc1740* and *vc2231*) in response to the availability of external fatty acids. Other sRNAs have recently been identified as part of so-called mixed regulatory circuits involving transcription factors and regulatory RNAs (68); however, none involved a 3'UTR-derived sRNA. Mixed FFLs come in two different designs with the sRNA working either as the top or middle regulator. However, only few mixed circuits



**Fig. 7.** Regulatory model for FarS-mediated fatty acid metabolism in *V. cholerae*. The *fabB* gene is part of the fatty acid biosynthesis regulon. It encodes  $\beta$ -ketoacyl-ACP synthase catalyzing the rate-limiting step in the synthesis of unsaturated fatty acid. The downstream reactions are catalyzed by FabG (3-ketoacyl-ACP reductase), FabZ (3-hydroxyacyl-ACP dehydratase), and FabV (enoyl-ACP reductase). FarS is produced from the 3'UTR of *fabB* and post-transcriptionally inhibits the expression of two paralogous *fadE* mRNAs. The *fadE* genes encode acyl-CoA dehydrogenase catalyzing the initial step in fatty acid  $\beta$ -oxidation. Here, long-chain fatty acids are transported across the outer and inner membranes by FadL and FadD, respectively. Following FadE activity, the remaining steps in fatty acid degradation are performed by a complex consisting of FadB and FadA.

have been analyzed for their regulatory dynamics. Two prime examples for sRNAs acting at the top and middle of the circuitry are RprA and Spot 42, respectively. RprA activates the production of *ppoS* and *ricI* at the posttranscriptional level, and RpoS is required for transcriptional activation of *ricI*. RicI inhibits plasmid conjugation in *Salmonella*, and, together, this system serves as a safety device to limit plasmid transfer under membrane-damaging conditions (69). The Spot 42 sRNA is repressed by the Crp transcriptional regulator and inhibits the expression of genes involved in the uptake and utilization of secondary carbon sources (constituting a type 4 coherent FFL). Spot 42 here modulates the dynamics of carbon utilization gene expression and reduces the overall leakiness of the system (70). Similarly, we discovered that FarS accelerates FadE repression when fatty acids are limited and serves as delay element when *V. cholerae* is transferred to high concentrations of fatty acids (Fig. 6 A–C). In addition, FarS inhibits FadE expression under regular growth conditions (Fig. 5 A and B), indicating a regulatory role when fatty acid concentrations are constant.

How this regulatory setup affects *V. cholerae*'s physiology is currently not fully understood, and, given various molecular mechanisms employed by Hfq-binding sRNAs (4), it is well possible that FarS regulates additional genes besides *vc1740* and *vc2231*. Nevertheless, it is interesting to note that *fabB* (which also produces FarS) constitutes the first gene in fatty acid biosynthesis, while the FarS target genes, the two *fadE* paralogs, are the first genes required for fatty acid degradation (Fig. 7). FabB carries out the rate-limiting step for the biosynthesis of unsaturated fatty acids (71) and has recently been employed to artificially control membrane viscosity in *E. coli* (72). Conversely, FadE, i.e., acyl-CoA dehydrogenase, is the rate-limiting enzyme for one cycle of oxidation of acyl-CoA (73). It is well conceivable that *V. cholerae* limits the production of fatty acid degradation genes when fatty acid biogenesis is activated, as high levels of FadE could result in a futile cycle in which newly synthesized fatty acids are degraded by the cellular machinery. At the transcriptional level, switching between fatty acid biosynthesis and degradation is controlled by FadR (29), and our data suggest that FarS improves the robustness of this system through a posttranscriptional control mechanism (Figs. 5 and 6). Such tight regulation of fatty metabolism might be particularly relevant for *V. cholerae*'s lifestyle. Transcriptomic analysis of *V. cholerae* infecting infant rabbits revealed a strong activation of fatty acid degradation genes (including *fadE*), which can be explained by the influx of long-chain fatty acids in the cecal fluid of infected animals (74). Indeed, recent work focusing on the role of CTX during the infection process showed that the acquisition of host-derived long-chain fatty acids is necessary for *V. cholerae*'s survival and replication in the host (17). Further, fatty acids also directly modulate the activity of the major virulence transcription factor ToxT, which is required for CTX production (18). We therefore conclude that fatty acid metabolism is a key feature of *V. cholerae*'s pathogenic lifestyle and possibly requires dynamic regulatory mechanisms, including the mixed feed-forward loop identified here, to balance fatty acid biosynthesis and degradation.

## Methods

**Bacterial Strains and Growth Conditions.** All strains used in this study are listed in *SI Appendix, Table S4*. Details on strain construction are provided in *SI Appendix, Materials and Methods*. *V. cholerae* and *E. coli* cells were grown under aerobic conditions in LB (Lennox broth) or M9 minimal medium (0.4% glucose, 0.4% casamino acids) at 37 °C unless stated otherwise. Where appropriate, antibiotics were used at the following concentrations: 100 µg/mL ampicillin, 20 µg/mL chloramphenicol, 50 µg/mL kanamycin, 50 U/mL polymyxin B, and 5,000 µg/mL streptomycin. When cultivated in minimal medium with fatty acids (sodium oleate; 0.005% [wt/vol] final concentration; Sigma; O3880) as sole carbon source, *V. cholerae* cells were inoculated ~1:1,000 from overnight cultures (M9 minimal medium with 0.4% glucose and 0.4% casamino acids) to the same starting OD<sub>600</sub> and grown for 10 h at 37 °C (200 rpm shaking

conditions). Serial dilutions were prepared and spotted on agar plates, and colony-forming units per milliliter were determined.

**Plasmids and DNA Oligonucleotides.** Plasmids and DNA oligonucleotides are listed in *SI Appendix, Tables S5 and S6*, respectively. Details on plasmid construction are provided in *SI Appendix, Materials and Methods*.

**Hfq Coimmunoprecipitation and cDNA Library Preparation.** *V. cholerae* wild-type (KPS-0014) and *hfq::3XFLAG*-tagged strains (KPS-0995) were cultivated in LB medium to low (OD<sub>600</sub> of 0.2) and high cell densities (OD<sub>600</sub> of 2.0). Cells equivalent to 50 OD<sub>600</sub> units were collected and subjected to coimmunoprecipitation as described previously (9), with slight modifications. Briefly, cells were resuspended in lysis buffer (20 mM Tris-HCl [pH 8], 150 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM DTT) and disrupted with 0.3-mL glass beads (Roth; 0.1 mm diameter) using a Bead Ruptor 4 (Omni). Cleared lysates were incubated with monoclonal anti-FLAG antibody (Sigma; F1804) and protein G Sepharose (Sigma; P3296). After stringent washing with lysis buffer, RNA and protein fractions were isolated by phenol-chloroform-isopropanol extraction. The RNA was subjected to DNase I (Thermo Fisher Scientific) digestion, and RNA integrity was confirmed using a Bioanalyzer (Agilent). cDNA libraries were prepared using the NEBNext Small RNA Library Prep Set for Illumina (NEB; E73005) according to the manufacturer's instructions.

**RIP-Seq Analysis.** cDNA libraries were sequenced on a HiSeq 1500 system in single-read mode with 100-nt read length. Demultiplexed raw reads were imported into CLC Genomics Workbench (Qiagen) and subjected to quality control and adaptor trimming. The trimmed reads were mapped to the *V. cholerae* reference genome [National Center for Biotechnology Information (NCBI) accession numbers NC\_002505.1 and NC\_002506.1, [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000006745.1](https://www.ncbi.nlm.nih.gov/assembly/GCF_000006745.1)] with standard parameter settings. sRNA annotations were added manually based on previously identified sRNA candidates (23). Fold enrichment in the *hfq::3XFLAG*-tagged samples over the untagged control samples was calculated using the CLC "Differential Expression for RNA-Seq" tool.

**Western Blot Analysis.** Western blot analysis of FLAG and GFP fusion proteins followed previously published protocols (24). Briefly, proteins were separated using SDS-PAGE and transferred to PVDF membranes. FLAG-tagged fusions were detected using anti-FLAG antibody (Sigma; F1804) and GFP-tagged fusions using anti-GFP antibody (Roche; no. 11814460001). RNAP served as loading control and was detected using anti-RNAP antibody (BioLegend; WP003). Signals were visualized on a Fusion FX imager (Vilber), and band intensities were quantified using the BIO-1D software (Vilber).

**RNA Isolation and Northern Blot Analysis.** Total RNA was prepared and transferred as described previously (75). Membranes were hybridized in Roti-Hybr-Quick buffer (Roth) at 42 °C with [<sup>32</sup>P] end-labeled DNA oligonucleotides or at 63 °C for riboprobes. Riboprobes were prepared using the MAXIscript T7 Transcription Kit (Thermo Fisher Scientific; AM1312). Signals were visualized on a Typhoon PhosphorImager (Amersham), and band intensities were quantified using the GelQuant software (BioChemLabSolutions). Oligonucleotides for Northern blot analyses are listed in *SI Appendix, Table S6*.

**Transcriptome Analysis.** *V. cholerae* wild-type cells harboring either pBAD-ctr or pBAD-*farS* were cultivated in triplicates to exponential phase (OD<sub>600</sub> of 0.5). Expression of FarS was induced by adding L-arabinose (0.2% final concentration). After 15 min of arabinose treatment, transcription was stopped by adding 0.2 volumes of stop mix (95% ethanol, 5% [vol/vol] phenol) and cells were harvested. Total RNA was prepared and subjected to Turbo DNase (Thermo Fisher Scientific) digestion. After confirming RNA integrity using a Bioanalyzer (Agilent), ribosomal RNA was depleted using the Ribo-Zero rRNA Removal Kit (Epicentre) for Gram-negative bacteria. cDNA libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB; E7760) according to the manufacturer's instructions. High-throughput sequencing was performed on a HiSeq 1500 system in single-read mode with 50-nt read length. Demultiplexed raw reads were trimmed for quality and adaptors and mapped to the *V. cholerae* reference genome (NCBI accession numbers NC\_002505.1 and NC\_002506.1) using CLC Genomics Workbench (Qiagen) with standard parameter settings. Reads mapped to annotated coding sequences were counted, and differential expression was calculated.

**Fatty Acid Transition Assays.** *V. cholerae* cells were grown to the desired cell densities in M9 minimal medium. To study the effect of addition of fatty acids, sodium oleate (0.005% [wt/vol] final concentration; Sigma; O3880) was added to the cultures, and RNA and protein samples were collected at

different time points (as indicated in figure legends). To analyze the reverse scenario, when fatty acids are removed, cultures were first cultivated in M9 minimal medium containing sodium oleate (0.005% [wt/vol] final concentration), washed at room temperature in 1× PBS, and resuspended for further growth in fresh M9 minimal medium lacking fatty acids. Again, RNA and protein samples were collected at the indicated time points. Expression of *FarS* was analyzed by Northern blots, and VC1740::3XFLAG and VC2231::3XFLAG protein levels were determined by Western blots.

**Data Availability.** The sequencing data of the RIP-seq experiment and the transcriptome analysis are available at the National Center for Biotechnology

Information Gene Expression Omnibus (GEO) database under the accession number GSE140516.

**ACKNOWLEDGMENTS.** We thank Helmut Blum, Stefan Krebs, Nikolai Peschek, and Andreas Brachmann for help with the RNA sequencing experiments and Andreas Starick for excellent technical support. We thank Jörg Vogel, Gisela Storz, and Mona Hoyos for comments on the manuscript and all members of the Papenfort lab for insightful discussions and suggestions. This work was supported by the GIF (Grant G-2411-416.13/2016), the Human Frontier Science Program (Grant CDA00024/2016-C), Deutsche Forschungsgemeinschaft (EXC 2051, Project 390713860), and the European Research Council (StG-758212).

1. E. Holmqvist, J. Vogel, RNA-binding proteins in bacteria. *Nat. Rev. Microbiol.* **16**, 601–615 (2018).
2. A. Santiago-Frangos, S. A. Woodson, Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip. Rev. RNA* **9**, e1475 (2018).
3. T. B. Updegrove, A. Zhang, G. Storz, Hfq: The flexible RNA matchmaker. *Curr. Opin. Microbiol.* **30**, 133–138 (2016).
4. K. Kavita, F. de Mets, S. Gottesman, New aspects of RNA-based regulation by Hfq and its partner sRNAs. *Curr. Opin. Microbiol.* **42**, 53–61 (2018).
5. D. J. Schu, A. Zhang, S. Gottesman, G. Storz, Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J.* **34**, 2557–2573 (2015).
6. A. Santiago-Frangos *et al.*, *Caulobacter crescentus* Hfq structure reveals a conserved mechanism of RNA annealing regulation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10978–10987 (2019).
7. I. Bilusic, N. Popitsch, P. Rescheneder, R. Schroeder, M. Lybecker, Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. *RNA Biol.* **11**, 641–654 (2014).
8. A. Zhang *et al.*, Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.* **50**, 1111–1124 (2003).
9. Y. Chao, K. Papenfort, R. Reinhardt, C. M. Sharma, J. Vogel, An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).
10. A. Sittka *et al.*, Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.* **4**, e1000163 (2008).
11. P. Sobrero, C. Valverde, The bacterial protein Hfq: Much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.* **38**, 276–299 (2012).
12. J. Vogel, B. F. Luisi, Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9**, 578–589 (2011).
13. Y. Chao, J. Vogel, The role of Hfq in bacterial pathogens. *Curr. Opin. Microbiol.* **13**, 24–33 (2010).
14. Y. Ding, B. M. Davis, M. K. Waldor, Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol. Microbiol.* **53**, 345–354 (2004).
15. A. J. Silva, J. A. Benitez, *Vibrio cholerae* biofilms and cholera pathogenesis. *PLoS Negl. Trop. Dis.* **10**, e0004330 (2016).
16. S. M. Butler, A. Camilli, Going against the grain: Chemotaxis and infection in *Vibrio cholerae*. *Nat. Rev. Microbiol.* **3**, 611–620 (2005).
17. F. Rivera-Chávez, J. J. Mekalanos, Cholera toxin promotes pathogen acquisition of host-derived nutrients. *Nature* **572**, 244–248 (2019).
18. M. J. Lowden *et al.*, Structure of *Vibrio cholerae* ToxT reveals a mechanism for fatty acid regulation of virulence genes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2860–2865 (2010).
19. E. S. Bradley, K. Bodi, A. M. Ismail, A. Camilli, A genome-wide approach to discovery of small RNAs involved in regulation of virulence in *Vibrio cholerae*. *PLoS Pathog.* **7**, e1002126 (2011).
20. B. W. Davies, R. W. Bogard, T. S. Young, J. J. Mekalanos, Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence. *Cell* **149**, 358–370 (2012).
21. Y. Shao, L. Feng, S. T. Rutherford, K. Papenfort, B. L. Bassler, Functional determinants of the quorum-sensing non-coding RNAs and their roles in target regulation. *EMBO J.* **32**, 2158–2171 (2013).
22. R. Herzog, N. Peschek, K. S. Fröhlich, K. Schumacher, K. Papenfort, Three autoinducer molecules act in concert to control virulence gene expression in *Vibrio cholerae*. *Nucleic Acids Res.* **47**, 3171–3183 (2019).
23. K. Papenfort, K. U. Förstner, J. P. Cong, C. M. Sharma, B. L. Bassler, Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E766–E775 (2015).
24. K. Papenfort *et al.*, A *Vibrio cholerae* autoinducer-receptor pair that controls biofilm formation. *Nat. Chem. Biol.* **13**, 551–557 (2017).
25. T. Song *et al.*, A new *Vibrio cholerae* sRNA modulates colonization and affects release of outer membrane vesicles. *Mol. Microbiol.* **70**, 100–111 (2008).
26. N. Peschek, M. Hoyos, R. Herzog, K. U. Förstner, K. Papenfort, A conserved RNA seed-pairing domain directs small RNA-mediated stress resistance in enterobacteria. *EMBO J.* **38**, e101650 (2019).
27. J. M. Liu *et al.*, Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.* **37**, e46 (2009).
28. M. Miyakoshi, Y. Chao, J. Vogel, Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Curr. Opin. Microbiol.* **24**, 132–139 (2015).
29. Y. Fujita, H. Matsuoka, K. Hirooka, Regulation of fatty acid metabolism in bacteria. *Mol. Microbiol.* **66**, 829–839 (2007).
30. J. W. Campbell, J. E. Cronan, Jr, The enigmatic *Escherichia coli* *fadE* gene is *yafH*. *J. Bacteriol.* **184**, 3759–3764 (2002).
31. M. Huber, K. Papenfort, Switching fatty acid metabolism by an RNA-controlled feed forward loop. National Center for Biotechnology Information Gene Expression Omnibus. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140516>. Deposited 17 November 2019.
32. J. F. Heidelberg *et al.*, DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477–483 (2000).
33. D. H. Lenz *et al.*, The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* **118**, 69–82 (2004).
34. B. M. Davis, M. K. Waldor, RNase E-dependent processing stabilizes MicX, a *Vibrio cholerae* sRNA. *Mol. Microbiol.* **65**, 373–385 (2007).
35. B. M. Davis, M. Quinones, J. Pratt, Y. Ding, M. K. Waldor, Characterization of the small untranslated RNA RyhB and its regulon in *Vibrio cholerae*. *J. Bacteriol.* **187**, 4005–4014 (2005).
36. S. Yamamoto *et al.*, Identification of a chitin-induced small RNA that regulates translation of the *tfoX* gene, encoding a positive regulator of natural competence in *Vibrio cholerae*. *J. Bacteriol.* **193**, 1953–1965 (2011).
37. A. L. Richard, J. H. Withey, S. Beyhan, F. Yildiz, V. J. DiRita, The *Vibrio cholerae* virulence regulatory cascade controls glucose uptake through activation of TarA, a small regulatory RNA. *Mol. Microbiol.* **78**, 1171–1181 (2010).
38. B. G. Sahagan, J. E. Dahlberg, A small, unstable RNA molecule of *Escherichia coli*: Spot 42 RNA. I. Nucleotide sequence analysis. *J. Mol. Biol.* **131**, 573–592 (1979).
39. G. A. Hansen *et al.*, Expression profiling reveals Spot 42 small RNA as a key regulator in the central metabolism of *Aliivibrio salmonicida*. *BMC Genomics* **13**, 37 (2012).
40. K. Papenfort *et al.*, SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Mol. Microbiol.* **62**, 1674–1688 (2006).
41. D. Dimastrogiovanni *et al.*, Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *eLife* **3**, e05375 (2014).
42. W. Shi *et al.*, The 40-residue insertion in *Vibrio cholerae* FadR facilitates binding of an additional fatty acyl-CoA ligand. *Nat. Commun.* **6**, 6032 (2015).
43. M. S. Guo *et al.*, MicL, a new  $\sigma$ E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev.* **28**, 1620–1634 (2014).
44. F. De Mets, L. Van Melderen, S. Gottesman, Regulation of acetate metabolism and coordination with the TCA cycle via a processed small RNA. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1043–1052 (2019).
45. M. Miyakoshi, G. Matera, K. Maki, Y. Sone, J. Vogel, Functional expansion of a TCA cycle operon mRNA by a 3' end-derived small RNA. *Nucleic Acids Res.* **47**, 2075–2088 (2019).
46. Y. Chao, J. Vogel, A 3' UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. *Mol. Cell* **61**, 352–363 (2016).
47. M. Grabowicz, D. Koren, T. J. Silhavy, The CpxQ sRNA negatively regulates Skp to prevent mistargeting of  $\beta$ -barrel outer membrane proteins into the cytoplasmic membrane. *MBio* **7**, e00312–e00316 (2016).
48. Y. Chao *et al.*, In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. *Mol. Cell* **65**, 39–51 (2017).
49. D. E. Cameron, J. M. Urbach, J. J. Mekalanos, A defined transposon mutant library and its use in identifying motility genes in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8736–8741 (2008).
50. M. Rehmsmeier, P. Steffen, M. Hochsmann, R. Giegerich, Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507–1517 (2004).
51. C. P. Corcoran *et al.*, Superfolder GFP reporters validate diverse mRNA targets of the classic porin regulator, MicF RNA. *Mol. Microbiol.* **84**, 428–445 (2012).
52. G. Kovacicova, W. Lin, R. K. Taylor, K. Skorupski, The fatty acid regulator FadR influences the expression of the virulence cascade in the El Tor biotype of *Vibrio cholerae* by modulating the levels of ToxT via two different mechanisms. *J. Bacteriol.* **199**, e00762–16 (2017).
53. L. Argaman *et al.*, Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.* **11**, 941–950 (2001).
54. J. Vogel *et al.*, RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* **31**, 6435–6443 (2003).
55. A. Smirnov *et al.*, Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11591–11596 (2016).
56. H. Ishikawa, H. Otaka, K. Maki, T. Morita, H. Aiba, The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3' poly(U) tail. *RNA* **18**, 1062–1074 (2012).
57. E. Holmqvist *et al.*, Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J.* **35**, 991–1011 (2016).
58. C. L. Kingsford, K. Ayanbule, S. L. Salzberg, Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.* **8**, R22 (2007).

59. T. B. Updegrove, A. B. Kouse, K. J. Bandyra, G. Storz, Stem-loops direct precise processing of 3' UTR-derived small RNA MicL. *Nucleic Acids Res.* **47**, 1482–1492 (2019).
60. T. B. Updegrove, S. A. Shabalina, G. Storz, How do base-pairing small RNAs evolve? *FEMS Microbiol. Rev.* **39**, 379–391 (2015).
61. D. Dar, R. Sorek, Extensive reshaping of bacterial operons by programmed mRNA decay. *PLoS Genet.* **14**, e1007354 (2018).
62. S. Melamed, P. P. Adams, A. Zhang, H. Zhang, G. Storz, RNA-RNA interactomes of ProQ and Hfq reveal overlapping and competing roles. *Mol. Cell* **77**, 411–425.e7 (2020).
63. A. J. Westermann *et al.*, The major RNA-binding protein ProQ impacts virulence gene expression in salmonella enterica serovar typhimurium. *MBio* **10**, e02504-18 (2019).
64. A. Smirnov, C. Wang, L. L. Drewry, J. Vogel, Molecular mechanism of mRNA repression in *trans* by a ProQ-dependent small RNA. *EMBO J.* **36**, 1029–1045 (2017).
65. E. Holmqvist, L. Li, T. Bischler, L. Barquist, J. Vogel, Global maps of ProQ binding in vivo reveal target recognition via RNA structure and stability control at mRNA 3' ends. *Mol. Cell* **70**, 971–982.e6 (2018).
66. H. M. Kim, J. H. Shin, Y. B. Cho, J. H. Roe, Inverse regulation of Fe- and Ni-containing SOD genes by a Fur family regulator Nur through small RNA processed from 3'UTR of the *sodF* mRNA. *Nucleic Acids Res.* **42**, 2003–2014 (2014).
67. Y. Malka *et al.*, Post-transcriptional 3'-UTR cleavage of mRNA transcripts generates thousands of stable uncapped autonomous RNA fragments. *Nat. Commun.* **8**, 2029 (2017).
68. M. Nitzan, R. Rehani, H. Margalit, Integration of bacterial small RNAs in regulatory networks. *Annu. Rev. Biophys.* **46**, 131–148 (2017).
69. K. Papenfort, E. Espinosa, J. Casadesús, J. Vogel, Small RNA-based feedforward loop with AND-gate logic regulates extrachromosomal DNA transfer in Salmonella. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4772–E4781 (2015).
70. C. L. Beisel, G. Storz, The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in Escherichia coli. *Mol. Cell* **41**, 286–297 (2011).
71. Y. Feng, J. E. Cronan, Escherichia coli unsaturated fatty acid synthesis: Complex transcription of the *fabA* gene and in vivo identification of the essential reaction catalyzed by FabB. *J. Biol. Chem.* **284**, 29526–29535 (2009).
72. I. Budin *et al.*, Viscous control of cellular respiration by membrane lipid composition. *Science* **362**, 1186–1189 (2018).
73. W. J. O'Brien, F. E. Frerman, Evidence for a complex of three beta-oxidation enzymes in Escherichia coli: Induction and localization. *J. Bacteriol.* **132**, 532–540 (1977).
74. A. Mandlik *et al.*, RNA-Seq-based monitoring of infection-linked changes in Vibrio cholerae gene expression. *Cell Host Microbe* **10**, 165–174 (2011).
75. K. S. Fröhlich, K. Haneke, K. Papenfort, J. Vogel, The target spectrum of SdsR small RNA in Salmonella. *Nucleic Acids Res.* **44**, 10406–10422 (2016).
76. F. Corpet, Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890 (1988).



## Supplementary Information

### Switching fatty acid metabolism by an RNA-controlled feed forward loop

Michaela Huber<sup>1,2</sup>, Kathrin S. Fröhlich<sup>1,2</sup>, Jessica Radmer<sup>2</sup>, and Kai Papenfort<sup>1,2,3 #</sup>

<sup>1</sup> Friedrich Schiller University, Institute of Microbiology, 07745 Jena, Germany

<sup>2</sup> Faculty of Biology I, Ludwig-Maximilians-University of Munich, 82152 Martinsried, Germany

<sup>3</sup> Microverse Cluster, Friedrich Schiller University Jena, 07743 Jena, Germany.

#### **This supplement contains:**

Figures S1 to S6

Supplementary Figure Legends

Supplementary Materials and Methods

Tables S1 to S6

Supplemental References

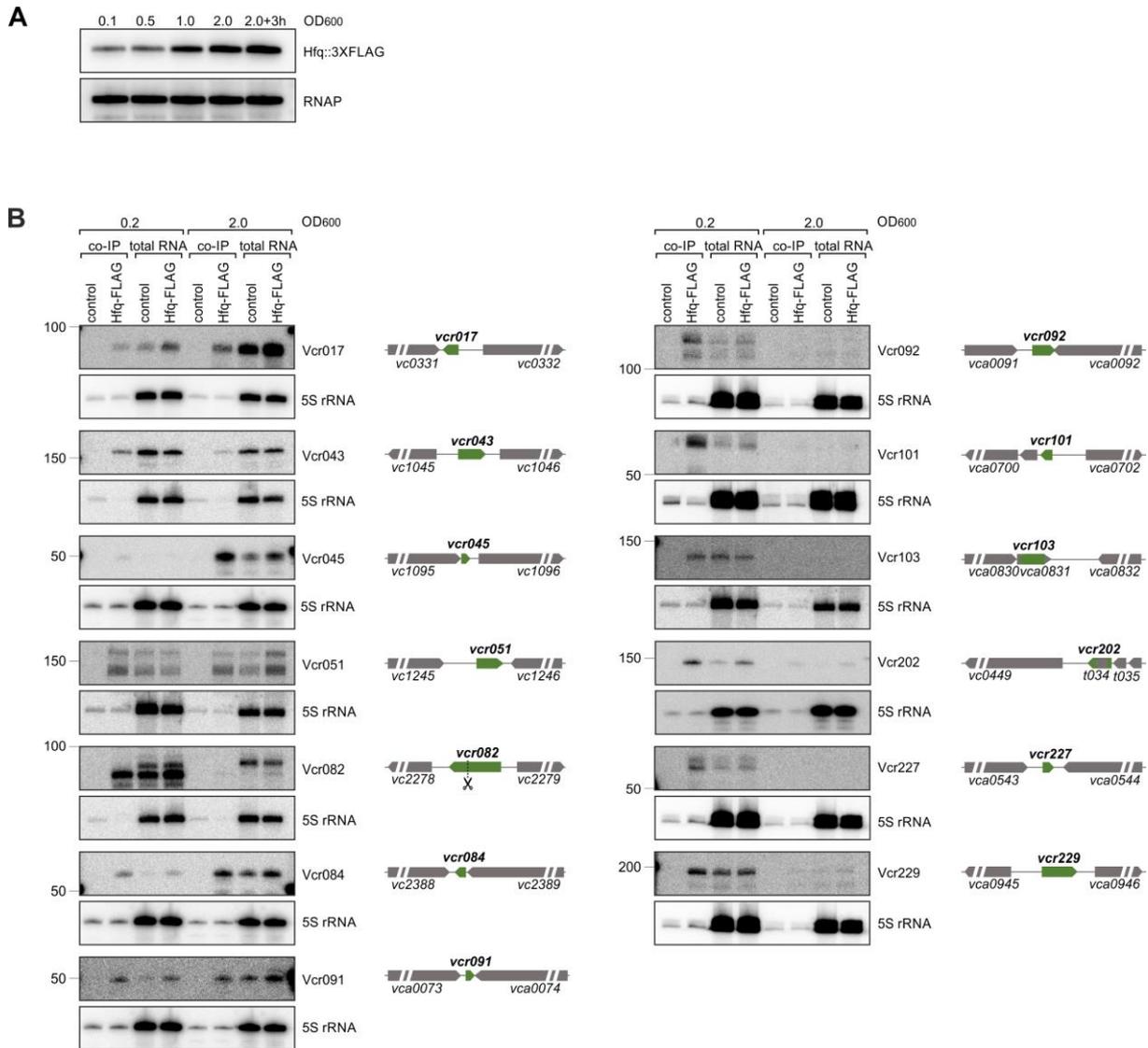
## TABLE OF CONTENTS

<b>Figure S1</b>	Expression of Hfq and further RIP-seq analysis
<b>Figure S2</b>	Hfq - sRNA binding experiments
<b>Figure S3</b>	Expression of <i>fabB</i> and <i>farS</i> requires FadR
<b>Figure S4</b>	Pulse induction of FarS and structure probing experiments
<b>Figure S5</b>	Hfq is required for FarS-mediated target regulation
<b>Figure S6</b>	Effect of FarS and fatty acids on FadE production

## Supplementary Materials and Methods

<b>Table S1</b>	Statistics of obtained and mapped cDNA reads for co-IP libraries
<b>Table S2</b>	Enriched transcripts in Hfq co-IP libraries
<b>Table S3</b>	Overview of new sRNA candidates
<b>Table S4</b>	Bacterial strains used in this study
<b>Table S5</b>	Plasmids used in this study
<b>Table S6</b>	DNA oligonucleotides used in this study

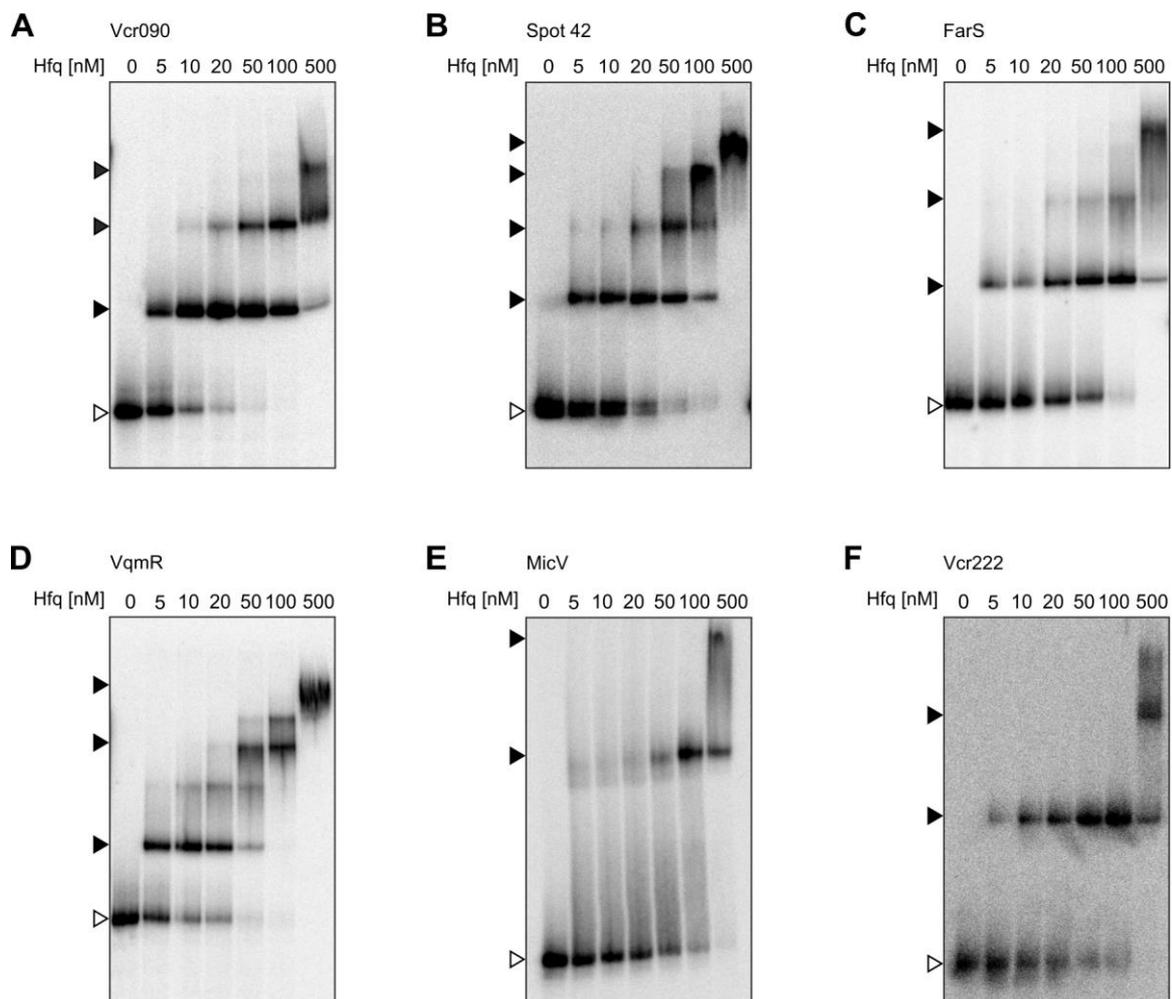
## Figure S1



### Figure S1: Expression of Hfq and further RIP-seq analysis

**A)** *V. cholerae* cells carrying a chromosomal 3XFLAG epitope at the *hfq* gene were cultivated in LB medium and protein samples were collected at the indicated OD<sub>600</sub> readings. Production of Hfq was monitored by Western blot analysis. RNAP served as loading control. **B)** Co-IP and total RNA (lysate) fractions were obtained from *V. cholerae* wild-type and *hfq*::3XFLAG strains following growth in LB medium to low (OD<sub>600</sub> of 0.2) and high cell densities (OD<sub>600</sub> of 2.0). The RNA was loaded on Northern blots and probed for the indicated sRNAs. 5S rRNA served as loading control. The genomic locations of the sRNAs are shown to the right. Flanking genes are shown in gray, sRNAs are shown in green. Scissors indicate putative processing site.

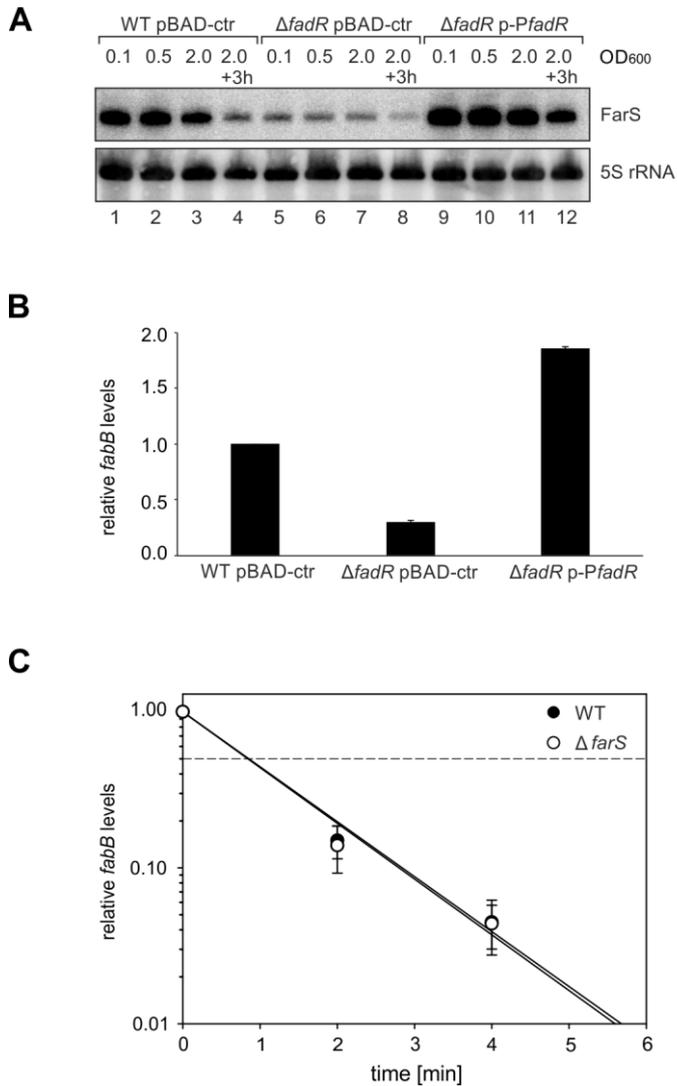
**Figure S2**



**Figure S2: Hfq - sRNA binding experiments**

**A – F)** Electrophoretic mobility shift assays (EMSAs) using *in vitro* synthesized, 5'-end-labelled sRNAs (4nM; A: Vcr090, B: Spot 42, C: FarS, D: VqmR, E: MicV, F: Vcr222) and increasing concentrations of purified *V. cholerae* Hfq protein. Open triangles indicate free sRNAs, solid triangles indicate sRNA-Hfq complexes.

**Figure S3**



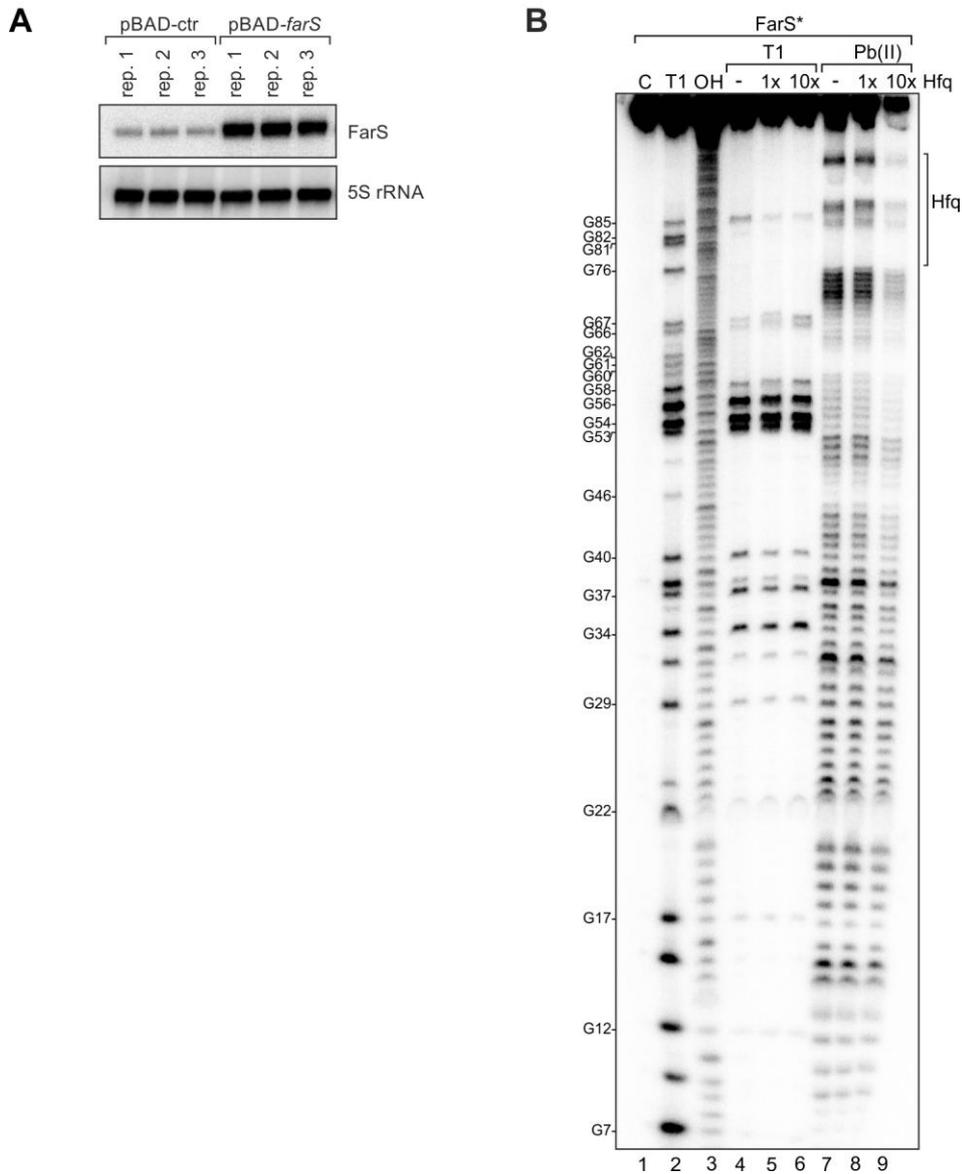
**Figure S3: Expression of *fabB* and *farS* requires FadR**

**A)** *V. cholerae* wild-type and  $\Delta$ *fadR* cells harboring either a control plasmid (pBAD-ctr) or a plasmid containing the *fadR* gene with its native promoter (p-*PfadR*) were cultivated in M9 minimal medium. Total RNA samples were collected at different stages of growth and Northern blot analysis was performed to determine FarS levels. 5S rRNA served as loading control.

**B)** *V. cholerae* wild-type and  $\Delta$ *fadR* strains harboring the indicated plasmids were cultivated in LB medium to an OD<sub>600</sub> of 1.0. RNA samples were collected and *fabB* mRNA levels were analyzed using qRT-PCR. Data are presented as mean  $\pm$  SD.

**C)** *V. cholerae* wild-type and  $\Delta$ *farS* strains were grown in LB medium to an OD<sub>600</sub> of 1.0. Cells were treated with rifampicin to terminate transcription. Total RNA samples were collected at the indicated time points and qRT-PCR was performed to monitor *fabB* transcript levels. Data are presented as mean  $\pm$  SD.

**Figure S4**

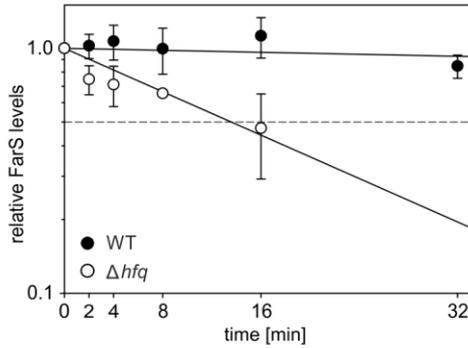
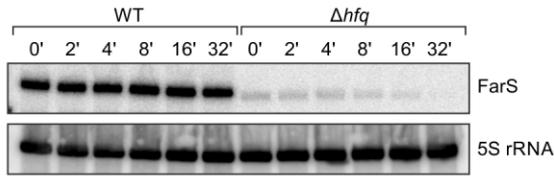


**Figure S4: Pulse induction of FarS and structure probing experiments**

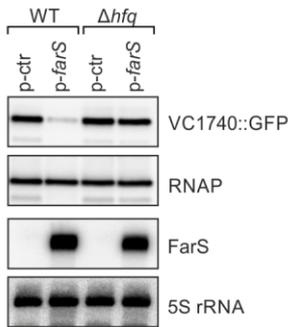
**A)** *V. cholerae* wild-type cells carrying the indicated plasmids were cultivated in LB medium to exponential phase ( $OD_{600}$  of 0.5) and induced with L-arabinose (0.2% final conc.) for 15 minutes. FarS levels were determined by Northern blot analysis and 5S rRNA was used as loading control. **B)** *In vitro* structure probing of 5'-end-labelled FarS sRNA (0.4 pmol) with RNase T1 (lanes 4 to 6) and lead(II) acetate (lanes 7 to 9) in the presence of 0.4 pmol (1x) or 4 pmol (10x) Hfq protein. RNase T1 and alkaline ladders of FarS were used to map the position of individual nucleotides. The positions of G residues are indicated relative to the transcriptional start site.

**Figure S5**

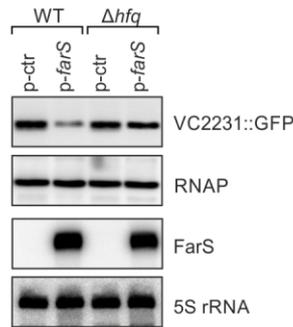
**A**



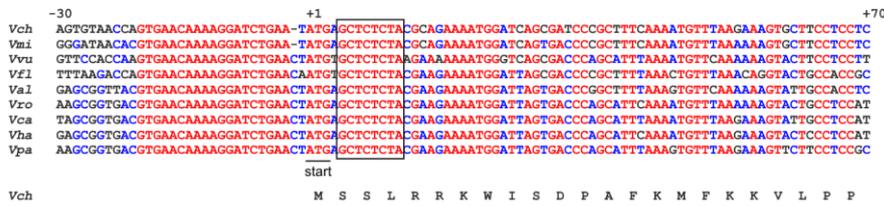
**B**



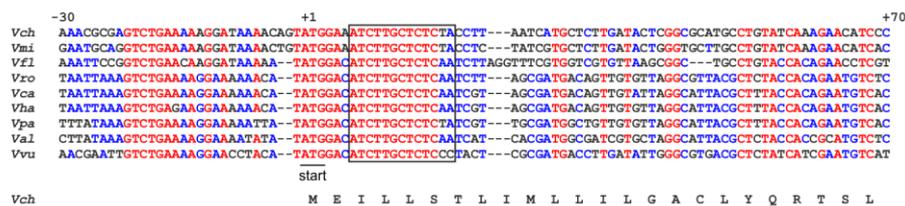
**C**



**D**



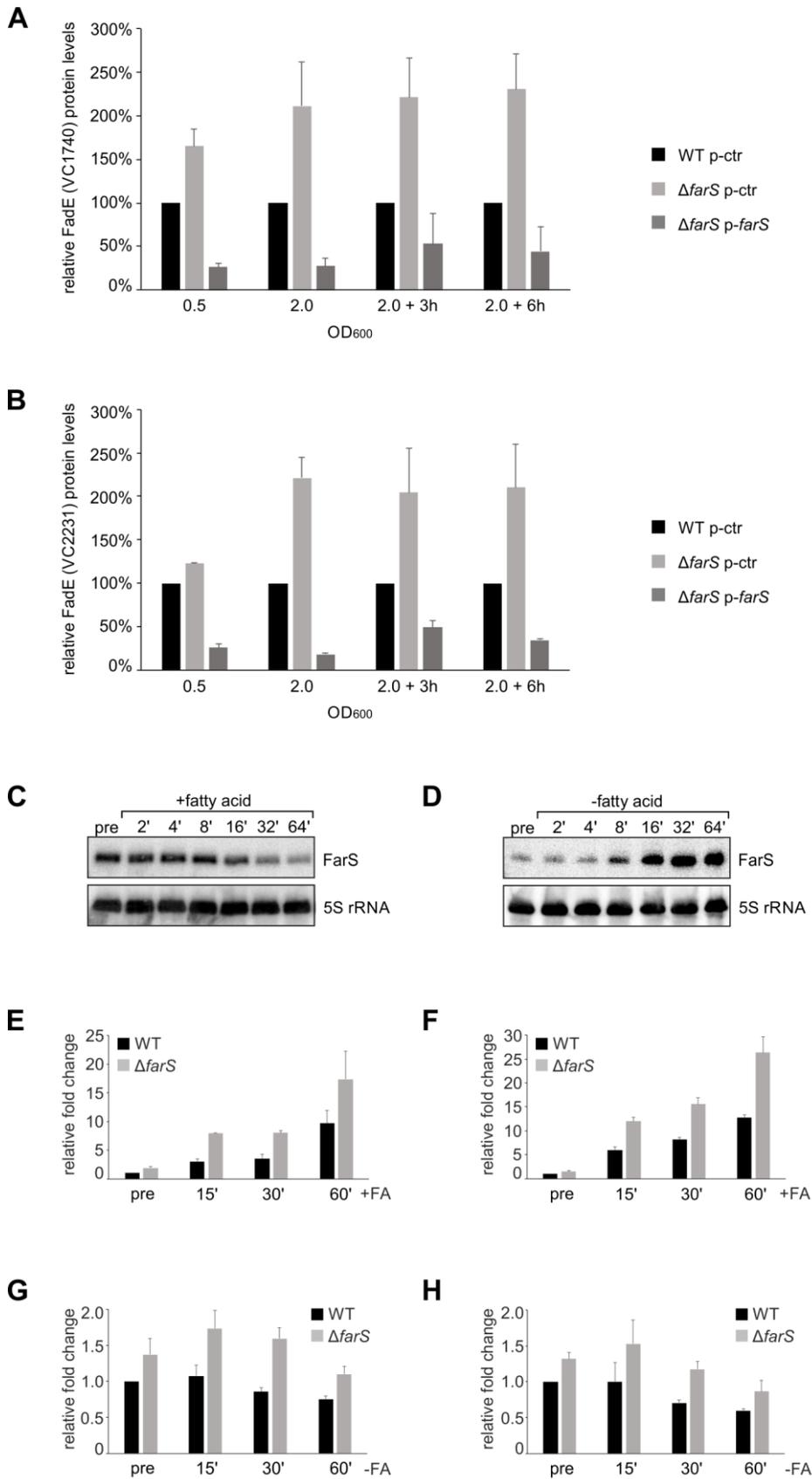
**E**



### Figure S5: Hfq is required for FarS-mediated target regulation

**A)** *V. cholerae* wild-type and  $\Delta hfq$  strains were cultivated in LB medium to an OD<sub>600</sub> of 1.5. Cells were treated with rifampicin and total RNA was collected at the indicated time points. Northern blot analysis was performed to monitor FarS levels. 5S rRNA was used as loading control. Data are presented as mean  $\pm$  SD. **B and C)** *E. coli* wild-type and  $\Delta hfq$  strains harboring a reporter plasmid for VC1740::GFP (A) or for VC2231::GFP (B) were co-transformed with plasmids p-ctr or p-*farS* and grown in LB medium to stationary phase (OD<sub>600</sub> of 2.0). GFP levels were analyzed by Western blotting and FarS levels were determined by Northern blot analysis. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. **D and E)** Alignments of *fadE* sequences in different *Vibrio* species (D: *vc1740*, E: *vc2231*). The sequences were aligned using the Multalign algorithm (1). Numbers above the sequences indicate the distance to the first nucleotides of the *fadE* start codons (marked with +1). The start codons are underlined. Black boxes indicate sequences base-pairing to FarS. The corresponding amino acid sequences for *V. cholerae* are shown below, respectively. *Vch*, *Vibrio cholerae*; *Vmi*, *Vibrio mimicus*; *Vvu*, *Vibrio vulnificus*; *Vfl*, *Vibrio fluvialis*; *Val*, *Vibrio alginolyticus*; *Vro*, *Vibrio rotiferianus*; *Vca*, *Vibrio campbellii*; *Vha*, *Vibrio harveyi*; *Vpa*, *Vibrio parahaemolyticus*.

**Figure S6**



### Figure S6: Effect of FarS and fatty acids on FadE production

**A and B)** *V. cholerae* wild-type and  $\Delta farS$  strains carrying a chromosomal 3XFLAG epitope either at the *vc1740* (A) or at the *vc2231* (B) gene and harboring the indicated plasmids were cultivated in M9 minimal medium. FadE::3XFLAG protein production (A: VC1740::3XFLAG, B: VC2231::3XFLAG) was analyzed by Western blotting (see Figs. 5A-B). Bar graphs show quantification of Western blots obtained from three independent biological replicates. FadE levels in the wild-type strains were set to 100%. Data are presented as mean  $\pm$  SD. **C)** *V. cholerae* wild-type cells were grown in M9 minimal medium and fatty acids (sodium oleate, 0.005% final conc.) were added when cells reached an OD<sub>600</sub> of 1.0. RNA samples were collected at the indicated time points before and after treatment. Northern blot analysis was performed to determine FarS levels. 5S rRNA was used as loading control. **D)** *V. cholerae* wild-type cells were cultivated in M9 minimal medium supplemented with external fatty acids (sodium oleate, 0.005% final conc.). When reaching an OD<sub>600</sub> of 1.0, cells were washed and transferred into fresh M9 minimal medium lacking fatty acids. RNA samples were collected at the indicated time points. The Northern blot was probed for FarS and 5S rRNA served as loading control. **E and F)** *V. cholerae* wild-type and  $\Delta farS$  strains carrying a chromosomal 3XFLAG epitope either at the *vc1740* (E) or at the *vc2231* (F) gene were cultivated in M9 minimal medium. When cells reached stationary phase (OD<sub>600</sub> of 2.0), fatty acids (+FA, sodium oleate, 0.005% final conc.) were added and FadE expression patterns were analyzed on Western blots (see Figs. 6B-C). Bar graphs show quantification of Western blots obtained from three independent biological replicates. Expression in the wild-type strain before treatment (pre) was set to 1. Data are presented as mean  $\pm$  SD. **G and H)** *V. cholerae* wild-type and  $\Delta farS$  strains carrying a chromosomal 3XFLAG epitope either at the *vc1740* (G) or at the *vc2231* (H) gene were cultivated in M9 minimal medium containing sodium oleate (0.005% final conc.). When reaching stationary phase, cells were washed and resuspended in M9 minimal medium lacking fatty acids (-FA) and FadE levels were monitored on Western blots (see Figs. 6D-E). Bar graphs show quantification of Western blots obtained from three independent biological replicates. Expression in the wild-type strain before treatment (pre) was set to 1. Data are presented as mean  $\pm$  SD.

## Supplementary Materials and Methods

### Plasmid construction

All plasmids used in this study are listed in Table S5, and all DNA oligonucleotides in Table S6. The plasmid pMH029 was constructed by amplifying the *hfq* gene (*vc0347*) from *V. cholerae* (KPS-0014) genomic DNA (gDNA) with KPO-2292 and KPO-2293 and inserting it into the linearized pTYB11 plasmid (NEB, KPO-2294/2295), using Gibson assembly (GA). To generate the sRNA expression plasmids pJR5 and pJR6, the *farS* gene was PCR amplified from gDNA, using primer sets KPO-2450/2452 and KPO-2451/2452, respectively. The fragments were fused to linearized pBAD1K (pMD004) or pEVS143 plasmid backbones (KPO-0196/1397 or KPO-0092/1397) via GA. pJR6 served as template to insert a single point mutation in the *farS* gene using site-directed mutagenesis and oligonucleotides KPO-3026/3027, yielding plasmid pJR14. The plasmid pMH034 was obtained by linearizing pMD004 with KPO-1792 and KPO-1397, and inserting the fragment amplified from gDNA with oligonucleotides KPO-2453 and KPO-2452, using GA. The *fabB* promoter truncation plasmids pJR8, pJR9 and pJR10 were cloned in the same way as pMH034 using the oligonucleotide combinations KPO-2454/2452, KPO-2455/2452 and KPO-2456/2452, respectively, for insert amplification. pMH034 served as template to construct pJR34 via site-directed mutagenesis using the oligonucleotides KPO-3963 and KPO-3964. To generate pJR22 by GA, the *fabB-farS* fragment was amplified from gDNA with KPO-3771 and KPO-2452 and pMD004 was linearized with pBAD-ATGrev and KPO-1397. To construct plasmid pJR12, the *farS* flanking regions were amplified with primer sets KPO-1278/2458 and KPO-2459/1281 respectively, and subsequently cloned via GA into the pKAS32 plasmid backbone, linearized with KPO-0267 and KPO-0268. pMH043 was obtained by linearizing pMH001 with KPO-1792 and KPO-1423 and inserting the fragment amplified from gDNA with oligonucleotides KPO-2764 and KPO-2765 via GA. GFP fusions were cloned as described previously (2). Briefly, *vc2231* (pMH037) and *vc1741/40* (pMH042) inserts for translational reporters were PCR amplified with the oligonucleotide sets KPO-2797/2798 and KPO-2546/2923 and introduced via GA into linearized pXG10 (KPO-1702/1703) and pXG30 (KPO-2662/1703) backbones, respectively. Single point mutations in the *vc2231* and *vc1740* genes were implemented by PCR using KPO-3030/3031 and KPO-3028/3029, resulting in plasmids pJR16 and pMH051, respectively. The plasmids pJR20 and pJR21 were constructed by GA using pKAS32 backbone that was linearized with KPO-0267 and KPO-0268. The insert fragments for pJR20 were amplified from KPS-0014 gDNA (KPO-3080/3081 and KPO-3084/3019) and KPS-0995 gDNA (KPO-3082/3083) that carries the 3XFLAG coding sequence. For pJR21, primer pairs KPO-3075/3076 (KPS-0014), KPO-3079/3015 (KPS-0014) and KPO-3077/3078 (KPS-0995) were used for insert amplification.

## Strain construction

All strains used in this study are listed in Table S4. *V. cholerae* C6706 was used as wild-type strain throughout the study. KPVC-11255 was constructed using natural transformation as described previously (3). Briefly, the flanking regions of *fadR* were amplified from *V. cholerae* gDNA with KPO-2766/2767 and KPO-2768/2769 and the FRT-flanked kanamycin cassette was amplified with KPO-1771/1772 from pBR-FRT-KAN-FRT (3). The three fragments were fused and amplified using KPO-2766/2769. Mutant cells were selected on kanamycin plates and confirmed by PCR using KPO-2698/1820. All other *V. cholerae* mutants were generated using the pKAS32 suicide vector (4) and established cloning strategies (5). Briefly, pKAS32-plasmids (pJR12, pMD003, pJR20 and pJR21) were conjugated into *V. cholerae* and cells were selected for ampicillin resistance. Polymyxin B was used to specifically inhibit *E. coli* growth. Single colonies were transferred to fresh plates and selected for streptomycin resistance. Mutants were confirmed by PCR and sequencing. KFS-01032 was established by P1 *vir* transduction of the  $\Delta hfq::KanR$  allele from the KEIO collection (6) using standard protocols.

## T7 transcription and 5' end labelling of RNA

DNA templates carrying a T7 promoter for *in vitro* synthesis of RNA were prepared by PCR using the oligonucleotides listed in Table S6. Template DNA (200 ng) was *in vitro* transcribed using the AmpliScribe T7-Flash transcription kit (Epicentre) following the manufacturer's recommendations. RNA size and integrity were verified on denaturing polyacrylamide gels. 5' end labelling was performed as described previously (7). Briefly, RNA (20 pmol) was dephosphorylated using 10 units of calf alkaline phosphatase (NEB), followed by P:C:I extraction and ethanol precipitation of RNA. Dephosphorylated RNA was incubated with [<sup>32</sup>P]- $\gamma$ ATP (20  $\mu$ Ci) and 1 unit of polynucleotide kinase (NEB) for 1 h at 37°C. Unincorporated nucleotides were removed using Microspin G-50 columns (GE Healthcare). Labelled RNA was loaded on a 6% / 7 M urea gel, cut from the gel, eluted overnight at 4°C with RNA elution buffer (0.1 M sodium acetate, 0.1% SDS, 10 mM EDTA), and recovered by P:C:I extraction.

## Purification of Hfq

The Hfq protein was expressed from the pTYB11 expression vector (NEB) in *E. coli* ER2566  $\Delta hfq$  cells and purified following the Impact Kit (NEB) protocol. Briefly, cells were grown to OD<sub>600</sub> of 0.5 and induced with IPTG (0.5 mM final conc.) for 15 h at 20°C. Cells were harvested, resuspended in column buffer (20 mM Tris-HCl [pH 8.5], 500 mM NaCl, 1 mM EDTA) and lysed by sonication. Cleared lysates were loaded on a column containing the chitin binding domain. After 40 h of incubation at room temperature, on column cleavage was induced using cleavage buffer (20 mM Tris-HCl [pH 8.5], 500 mM NaCl, 50 mM DTT, 1mM EDTA). Protein purification

was verified by SDS-PAGE analysis. The Hfq protein was concentrated and buffer was exchanged to storage buffer (25 mM Tris-HCl [pH 7.6], 150 mM NaCl, 0.5 mM EDTA) using 5 kDA MWCO Vivaspin columns (GE Healthcare).

### **Electrophoretic Mobility Shift Assays (EMSA)**

To analyze complex formation between sRNAs and Hfq *in vitro*, gel shift assays were performed following previously established protocols (8). Briefly, 5' end-labelled RNA (4 pmol) was supplemented with 1x structure buffer (0.01 M Tris-HCl [pH 7], 0.1 M KCl, 0.01 M MgCl<sub>2</sub>) and 1 µg yeast RNA and incubated with increasing concentrations of purified Hfq or Hfq dilution buffer (1x structure buffer, 1% [vol/vol] glycerol, 0.1% [vol/vol] Triton X-100) at 37°C for 15 min. Prior to loading, reactions were mixed with native loading buffer (50% glycerol, 0.5x TBE, 0.2% [wt/vol] bromphenol blue) and separated by native PAGE. Signals were visualized on a Typhoon Phosphorimager (Amersham).

### **Quantitative real-time PCR (qRT-PCR)**

Quantitative real-time PCR was performed as described previously (9). Briefly, RNA was extracted using the SV total RNA Isolation System (Promega) and *fabB* transcript levels were measured using the Luna Universal One-Step RT-qPCR Kit (NEB) and the MyiQ Single-Color Real-Time PCR Detection System (Bio-Rad). Oligonucleotides used for qRT-PCR are listed in Table S6.

### **Transcript stability experiments**

In order to analyze RNA stability, cells were treated with rifampicin (250 µg/ml final conc.) at the designated ODs to terminate transcription. RNA samples were collected at the indicated time points and transcript levels were determined either by Northern blot analysis or by qRT-PCR.

### **RNA structure probing**

RNA structure probing was carried out as described previously (10) with few modifications. In brief, 0.4 pmol 5' end-labelled FarS sRNA was denatured, quickly chilled on ice and mixed with 0.4 pmol or 4 pmol of purified *V. cholerae* Hfq protein or an equal volume of Hfq dilution buffer in the presence of 1x structure buffer and 1 µg yeast RNA. Samples were incubated at 37°C for 15 min, and treated with RNase T1 (0.1 U; Ambion, #AM2283) for 2.5 min or with lead(II) acetate (5 mM final conc.; Sigma, #316512) for 1.5 min. Reactions were stopped by the addition of 2 vol. stop/precipitation buffer (1 M guanidinium thiocyanate, 0.167% N-laurylsarcosine, 10 mM DTT, 83% 2-propanol). RNA was precipitated for 2 h at -20°C, and collected

by centrifugation (30 min, 4°C, 13.000 rpm). Samples were dissolved in GLII loading buffer, and separated on 10% polyacrylamide sequencing gels.

**Table S1: Statistics of obtained and mapped cDNA reads for co-IP libraries**

strain	condition	replicate	number of reads (million)	reads mapped on chr I (million)	reads mapped on chr II (million)	mapped reads in total
untagged control	0.2	I	11.0	10.4	0.3	97.6%
untagged control	0.2	II	13.3	12.5	0.4	97.8%
Hfq::3XFLAG	0.2	I	13.5	12.5	0.7	97.9%
Hfq::3XFLAG	0.2	II	12.3	10.8	1.2	97.6%
untagged control	2.0	I	39.1	36.9	0.9	96.7%
untagged control	2.0	II	12.1	11.7	0.2	98.0%
Hfq::3XFLAG	2.0	I	7.9	6.8	0.8	96.7%
Hfq::3XFLAG	2.0	II	14.5	12.6	1.6	97.5%

**Table S3: Overview of new sRNA candidates**

name	start	stop	orientation	size (nt)	enriched in Hfq co-IP
<b>chromosome 1</b>					
Vcr200	218.072	218.336	sense	265	yes
Vcr201	455.266	455.354	sense	89	yes
Vcr202	481.285	481.138	antisense	148	yes
Vcr203	606.829	606.883	sense	55	yes
Vcr204	677.937	678.072	sense	136	yes
Vcr205	714.030	714.121	sense	92	no
Vcr206	944.382	944.313	antisense	70	no
Vcr207	1.106.591	1.106.734	sense	144	yes
Vcr208	1.531.755	1.531.675	antisense	81	yes
Vcr209	1.578.023	1.578.082	sense	60	yes
Vcr210	1.582.874	1.582.933	sense	60	yes
Vcr211	1.861.483	1.861.570	sense	88	no
Vcr212	2.000.800	2.001.130	sense	331	yes
Vcr213	2.059.854	2.060.038	sense	185	yes
Vcr214	2.376.028	2.376.142	sense	115	yes
Vcr215	2.396.723	2.396.633	antisense	91	yes
Vcr216	2.518.934	2.518.785	antisense	150	yes
Vcr217	2.537.176	2.537.226	sense	51	yes
Vcr218	2.558.990	2.558.878	antisense	113	yes
Vcr219	2.639.102	2.639.035	antisense	68	yes
Vcr220	2.653.872	2.654.007	sense	136	yes
Vcr221	2.669.966	2.670.113	sense	148	no
Vcr222	2.783.908	2.783.840	antisense	69	yes
Vcr223	2.855.213	2.855.093	antisense	121	no
<b>chromosome 2</b>					
Vcr224	14.695	14.764	sense	70	yes
Vcr225	42.166	42.215	sense	50	yes
Vcr226	334.397	334.515	sense	119	no
Vcr227	479.958	480.023	sense	66	yes
Vcr228	787.266	787.431	sense	166	yes
Vcr229	897.527	897.726	sense	200	yes
Vcr230	937.994	938.066	sense	73	yes

**Table S4: Bacterial strains used in this study**

Strain	Relevant markers / genotype	Reference / Source
<b><i>V. cholerae</i></b>		
KPS-0014	C6706 wild-type	(11)
KPS-0054	C6706 $\Delta hfq$	(12)
KPS-0995	C6706 $hfq::3Xflag$	(13)
KPVC-11063	C6706 $\Delta farS$	This study
KPVC-11255	C6706 $\Delta fadR::kan$	This study
KPVC-11437	C6706 $\Delta vc2231/\Delta vc1740$	This study
KPVC-11488	C6706 $\Delta farS vc2231::3Xflag$	This study
KPVC-11492	C6706 $vc2231::3Xflag$	This study
KPVC-11525	C6706 $vc1740::3Xflag$	This study
KPVC-11526	C6706 $\Delta farS vc1740::3Xflag$	This study
KPVC-11527	C6706 $mre-3071 \Delta farS$	This study
<b><i>E. coli</i></b>		
Top10	<i>F- mcrA</i> $\Delta(mrr-hsdRMS-mcrBC)$ $\phi 80lacZ\Delta M15 \Delta lacX74 nupG recA1 araD139 \Delta(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 \lambda-$	Invitrogen
S17 $\lambda$ pir	$\Delta lacU169 (\Phi lacZ\Delta M15), recA1, endA1, hsdR17, thi-1, gyrA96, relA1, \lambda pir$	(14)
ER2566	<i>fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 <math>\Delta(mcrC-mrr) 114::IS10</math></i>	New England Biolabs
KFS-01032	ER2566 $\Delta hfq::kan$	This study
KPEC-50812	MC4100 $\Delta hfq$	(15)

**Table S5: Plasmids used in this study**

Plasmid trivial name	Plasmid stock name	Relevant fragment	Comment	Origin, marker	Reference
p-ctr	pCMW-1		control plasmid	P15A, Kan <sup>R</sup>	Papenfort plasmid collection
pBAD1K-ctr	pMD004		control plasmid	P15A, Kan <sup>R</sup>	Papenfort plasmid collection
pBAD1C-ctr	pMH001		control plasmid	P15A, Cm <sup>R</sup>	Papenfort plasmid collection
pEVS143	pEVS143	<i>Ptac</i> promoter	constitutive overexpression plasmid	P15A, Kan <sup>R</sup>	Papenfort plasmid collection
pKAS32	pKAS32		suicide plasmid for allelic exchange	R6K, Amp <sup>R</sup>	(4)
pXG10- <i>gfp</i>	pXG10- <i>gfp</i>	<i>lacZ::gfp</i>	template plasmid for translational reporters	pSC101*, Cm <sup>R</sup>	(2)
pXG30- <i>gfp</i>	pXG30- <i>gfp</i>	<i>flag::lacZ::gfp</i>	template plasmid for translational reporters	pSC101*, Cm <sup>R</sup>	(2)
pTYB11- <i>hfq</i>	pMH029	<i>hfq</i> ( <i>vc0347</i> )	intein fusion vector for Hfq protein purification	pBR322, Amp <sup>R</sup>	This study
pBAD- <i>farS</i>	pJR5	<i>farS</i>	<i>farS</i> expression plasmid	P15A, Kan <sup>R</sup>	This study
p- <i>farS</i>	pJR6	<i>farS</i>	<i>farS</i> expression plasmid	P15A, Kan <sup>R</sup>	This study
p- <i>PfabB-farS</i>	pMH034	<i>PfabB-farS</i>	<i>fabB-farS</i> expression plasmid	P15A, Kan <sup>R</sup>	This study
p- <i>PfabB-farS</i> -300 bp	pJR8	<i>fabB-farS</i>	<i>fabB-farS</i> promoter truncation plasmid	P15A, Kan <sup>R</sup>	This study
p- <i>PfabB-farS</i> -600 bp	pJR9	<i>fabB-farS</i>	<i>fabB-farS</i> promoter truncation plasmid	P15A, Kan <sup>R</sup>	This study
p- <i>PfabB-farS</i> -900 bp	pJR10	<i>fabB-farS</i>	<i>fabB-farS</i> promoter truncation plasmid	P15A, Kan <sup>R</sup>	This study
pKAS32- $\Delta$ <i>farS</i>	pJR12	up/downstream flanks of <i>farS</i>	suicide plasmid for <i>farS</i> knock-out	R6K, Amp <sup>R</sup>	This study
p- <i>PfadR</i>	pMH043	<i>PfadR</i>	<i>fadR</i> expression plasmid	P15A, Cm <sup>R</sup>	This study
p- <i>PfabB-farS</i> > GGG	pJR34	<i>PfabB-farS</i>	mutated <i>rne</i> site (TTT1-3GGG)	P15A, Kan <sup>R</sup>	This study
pKAS- <i>rne</i> -3071	pMD003	<i>rne</i> -3071	suicide plasmid for <i>rne</i> C202T base mutation	R6K, Amp <sup>R</sup>	Papenfort plasmid collection
pBAD- <i>fabB-farS</i>	pJR22	<i>fabB-farS</i>	<i>fabB-farS</i> expression plasmid	P15A, Kan <sup>R</sup>	This study
p- <i>farS</i> *	pJR14	<i>farS</i> *	<i>farS</i> * (G54C) expression plasmid	P15A, Kan <sup>R</sup>	This study
pXG10- <i>vc2231</i>	pMH037	<i>vc2231::gfp</i>	translational reporter for <i>vc2231</i>	pSC101*, Cm <sup>R</sup>	This study
pXG10- <i>vc2231</i> *	pJR16	<i>vc2231*::gfp</i>	translational reporter for <i>vc2231</i> * (C17G)	pSC101*, Cm <sup>R</sup>	This study
pXG30-1741/40	pMH042	<i>flag::vc1741 vc1740::gfp</i>	translational reporter for <i>vc1741</i> and <i>1740</i>	pSC101*, Cm <sup>R</sup>	This study
pXG30- <i>vc1741/40</i> *	pMH051	<i>flag::vc1741 vc1740*::gfp</i>	translational reporter for <i>vc1741</i> and <i>1740</i> * (C10G)	pSC101*, Cm <sup>R</sup>	This study
pKAS32- <i>vc2231::3Xflag</i>	pJR20	<i>vc2231::3Xflag</i>	<i>vc2231::3Xflag</i> allelic replacement	R6K, Amp <sup>R</sup>	This study
pKAS32- <i>vc1740::3Xflag</i>	pJR21	<i>vc1740::3Xflag</i>	<i>vc1740::3Xflag</i> allelic replacement	R6K, Amp <sup>R</sup>	This study

**Table S6: DNA oligonucleotides used in this study**

Name	Sequence 5' to 3'	Description
KPO-0009	CTACGGCGTTTCACTTCTGAGTTC	<i>E.c.</i> 5S oligoprobe
KPO-0063	CGTCTATAAGTGTGAACAATGGTG	Qrr4 oligoprobe
KPO-0092	CCACACATTATACGAGCCGA	plasmid construction
KPO-0196	GGAGAAACAGTAGAGAGTTGCG	plasmid construction
KPO-0216	AGTCGAGGACTCAGTTTATGATTA	Vcr017 oligoprobe
KPO-0243	TTCGTTTCACTTCTGAGTTCCG	<i>V.ch.</i> 5S oligoprobe
KPO-0267	TAATAGGCCTAGGATGCATATG	plasmid construction
KPO-0268	CGTTAACCAACCGGTACCTCTA	plasmid construction
KPO-0331	GAGCCAATCTACAATTCATCAGA	Vcr090 oligoprobe
KPO-0452	ATCTTGTGACGTGTAGAAGAGGTT	VqmR oligoprobe
KPO-0513	GTTTTTTTTAATACGACTCACTATAGCAGAGCATGAGTTGCATGAC	VqmR T7 transcription
KPO-0514	AAAAAAGCCAGCCTGAAGACG	VqmR T7 transcription
KPO-0820	GGCCTTCTTAGAGTCTTCTAAGAA	MicV oligoprobe
KPO-0821	AGGTTGTCAGAGAGGCCTTGA	Vcr084 oligoprobe
KPO-0822	GCCAGGTGAATAATGCGCTTG	Vcr092 oligoprobe
KPO-0842	GTAAGCAATTAACCTACGCCAATTG	Vcr043 oligoprobe
KPO-0845	TTGGCCCGTCACAGGCTGAA	Vcr045 oligoprobe
KPO-0873	CTCTCCATGGGACAGAGTCT	FarS oligoprobe
KPO-1278	TAGAGGTACCGTTGTTAACGCACCGGTATGGGTATTATTTCCG	plasmid construction
KPO-1281	CATATGCATCCTAGGCCTATTAGTTGGCTCATCACATACCTC	plasmid construction
KPO-1397	GATCCGGTGATTGATTGAGC	plasmid construction
KPO-1423	TCTAGATTAATCAGAACGCAGAAG	plasmid construction
KPO-1702	ATGCATGTGCTCAGTATCTCTATC	plasmid construction
KPO-1703	GCTAGCGGATCCGCTGG	plasmid construction
KPO-1771	TAATGTCCGAGTAGGCTGGAGCTGCTTCAAGTTCC	strain construction
KPO-1772	CTTCCAGAGACATATGAATATCCTCCTTAGTTCCTATTTC	strain construction
KPO-1792	CAGTGCGCCCTTTTATAGTC	plasmid construction
KPO-1820	ATGGATACTTTCTCGGCAG	strain construction
KPO-2002	GTTTACCATCGCTTATAGTTATA	Vcr091 oligoprobe
KPO-2010	TAAAGCTTTCAACCTGTGACG	Vcr222 oligoprobe
KPO-2075	GTGTCTATGGCACAACCTTTTAA	Vcr202 oligoprobe
KPO-2077	CCGCGAAAAGTAGGTTGTTTT	Vcr229 oligoprobe
KPO-2155	GGTATCTAAATCTTTTCGATACG	Vcr227 oligoprobe
KPO-2178	GTTTTTTAATACGACTCACTATAGGGAGGGTGAATCATATCGACCAAATTTG	Vcr082 riboprobe
KPO-2179	GTCTGCAATGTTCTGGAACC	Vcr082 riboprobe
KPO-2292	CCCAGTTTGTGTACAGAACATGGCTAAGGGGCAATCTCTA	plasmid construction
KPO-2293	CGGATCCCCTTCTGCAGTTACTCTTCAGACTTCTCTGC	plasmid construction
KPO-2294	GTTCTGTACAACAACCTGGG	plasmid construction
KPO-2295	CTGCAGGAAGGGGATCCG	plasmid construction
KPO-2378	GGTAACCCAGAACTACCACTG	<i>recA</i> qRT-PCR
KPO-2379	CACCACTTCTTCGCCTTCTT	<i>recA</i> qRT-PCR
KPO-2450	CGCAACTCTACTGTTTTCTCCTTTCCAGAACAGATTAGTTTTCGC	plasmid construction
KPO-2451	TCGGCTCGTATAATGTGTGGTTTTCCAGAACAGATTAGTTTTCGC	plasmid construction
KPO-2452	GCTCAATCAATCACCGGATCCAATCAAAGTTGCAGGCATTG	plasmid construction
KPO-2453	GACTATAAAAAGGCGCACTGCAGACCATTGACGTTAGAGAAA	plasmid construction
KPO-2454	GACTATAAAAAGGCGCACTGCTAAAAGCTGGCAAGTCAGG	plasmid construction
KPO-2455	GACTATAAAAAGGCGCACTGGTGGTCTTACATGGTG	plasmid construction
KPO-2456	GACTATAAAAAGGCGCACTGGGTTTTCGTGATCTCTGGCG	plasmid construction
KPO-2458	CTAGGCCGCCGGGCAAACCTGTGTTGGATCTGGTGCG	plasmid construction
KPO-2459	TTTGCCCGCGGCCTAG	plasmid construction
KPO-2546	CCAGCGGATCCGCTAGCCACTTTCTTAAACATTTTGAAAGC	plasmid construction
KPO-2549	AAAAATACCCGACGACCTAGG	FarS T7 transcription
KPO-2650	CCCTCTTAGGAAAAATTGTCAC	Vcr101 riboprobe
KPO-2651	GTTTTTTAATACGACTCACTATAGGGAGGCACCATAAAAAAGCCCCG	Vcr101 riboprobe
KPO-2662	TTTATCGTCGTCATCTTTGTAG	plasmid construction
KPO-2698	GTTTCGCAGCTACCACTGG	strain construction

KPO-2764	GACTATAAAAAGGCGCACTGGATAGTGTGAGCTGTGTCC	plasmid construction
KPO-2765	CTGCGTTCTGATTTAATCTAGATTAGCAATCGTCTTCAGTAAAATTG	plasmid construction
KPO-2766	GGCAATAACGATACTCAAGTTC	strain construction
KPO-2767	TCCAGCCTACTCCGACATTATCTAGCACTGTTTCGTTTTTCGTTA	strain construction
KPO-2768	TATTCATATGTCTCTGGAAGCCACTAGTTGGTGTACGTCCG	strain construction
KPO-2769	GCTATCGAAAGGAGAACTTTGG	strain construction
KPO-2797	GAGATACTGAGCACATGCAT ATCCGAACCCGCGCGCTT	plasmid construction
KPO-2798	GAGCCAGCGGATCCGCTAGTTGATACAGGCATGCGCCG	plasmid construction
KPO-2923	CTACAAAGATGACGACGATAAATCGTTAAGCGAATTGCGCCC	plasmid construction
KPO-3015	CATATGCATCCTAGGCCTATTAGAGCGGCATCACAGGAATC	plasmid construction
KPO-3019	CATATGCATCCTAGGCCTATTACTCCAATCTACAACCTACGAC	plasmid construction
KPO-3026	GTCCCATGCAGAGCGGGATAGGATCCTT	plasmid construction
KPO-3027	CTCTGCATGGGACAGAGTCTGCGTCTG	plasmid construction
KPO-3028	GAGCTCTGTACGCAGAAAATGGATCAGCG	plasmid construction
KPO-3029	CATTTTCTGCGTACAGAGCTCATATTCAG	plasmid construction
KPO-3030	CTTGCTCTGTACCTTAATCATGCTCTTG	plasmid construction
KPO-3031	GATTAAGGTACAGAGCAAGATTTCCATAC	plasmid construction
KPO-3075	TAGAGGTACCGGTTGTTAACGGATGGACCAATGAACATCTGG	plasmid construction
KPO-3076	TGCCCACTGTTAACTTAGG	plasmid construction
KPO-3077	CCTAAGTTAAACAGTGTGGCAGACTACAAAGACCATGACGG	plasmid construction
KPO-3078	CCTCGATACTCTATTTTATTTGTTATTACTATTTATCGTCATCTTTGTAG	plasmid construction
KPO-3079	TAACAAATAAAATAGAGTATCGAGG	plasmid construction
KPO-3080	TAGAGGTACCGGTTGTTAACGGTAAAGGCATCTGTTTAGGCC	plasmid construction
KPO-3081	AGCCACTTCAGCTTTACGTTG	plasmid construction
KPO-3082	CAACGTAAGCTGAAGTGGCTGACTACAAAGACCATGACGG	plasmid construction
KPO-3083	GAATACTGCTCAATGTGGAACCTATTACTATTTATCGTCATCTTTGTAG	plasmid construction
KPO-3084	TAAGTCCACATTGAGCAGTATTC	plasmid construction
KPO-3139	GGACAGAGTCTGCGTCTG	oligoprobe for mutated FarS (G54C)
KPO-3387	GTAACGCGGTTGAGCTTAT	<i>fabB</i> qRT-PCR
KPO-3388	CATGGTTTGTGACCAGTAGAG	<i>fabB</i> qRT-PCR
KPO-3726	CAGCCTAATCCAATAACGTGAAAC	Spot 42 oligoprobe
KPO-3771	GCTAACAGGAGGAATTAACCATGAAACGAGTCGTCATCAC	plasmid construction
KPO-3794	GACCCTTTCTTTGTTGCTC	Vcr103 oligoprobe
KPO-3751	ACCTGATTCCATCCCGAA	5S qRT-PCR
KPO-3752	TGGCGATGTTCTACTCTCA	5S qRT-PCR
KPO-3963	CAACACAGGGGCCAGAACAGATTAGTTTCGC	plasmid construction
KPO-3964	CTGTTCTGGCCCTGTGTTGGATCTGGTG	plasmid construction
KPO-4131	AAGAAAAAGCCCTAACCTAGTAC	MicV T7 transcription
KPO-4154	GTTTTTTTTAATACGACTCACTATAGACCACTGTTTTTCTTAGAAGAC	MicV T7 transcription
KPO-4249	GTTTTTTTTAATACGACTCACTATAGGTTTCCAGAACAGATTAGTTTCGC	FarS T7 transcription
KPO-5083	GTTTTTTTTAATACGACTCACTATAGGAATAGACAACCTTTTGCCT	Vcr090 T7 transcription
KPO-5084	AAAAAAGAGCGAGCTATTTAAAC	Vcr090 T7 transcription
KPO-5085	GTTTTTTTTAATACGACTCACTATAGGACTCTAATCATAATTTATTTTGTGC	Vcr222 T7 transcription
KPO-5086	AGCTTTCAACCTGTGACGAA	Vcr222 T7 transcription
KPO-5087	GTTTTTTTTAATACGACTCACTATAGGCGTAGGGTACAGAGGTAAG	Spot 42 T7 transcription
KPO-5088	AATAAAAAACGCCCCAGTCAAA	Spot 42 T7 transcription
KPO-5415	CAACGGGAGAGAAAACGGTT	VSSrna24 oligoprobe
KPO-5468	CTTGTTAGGCTCATCACTCTTC	Vcr051 riboprobe
KPO-5469	GTTTTTTTTAATACGACTCACTATAGGGAGGCAGTTCAGCACAACTCAATAC	Vcr051 riboprobe
pBAD-ATGrev	GGTTAATTCCTCCTGTTAGC	plasmid construction

## Supplemental References

1. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16(22):10881-10890.
2. Corcoran CP, *et al.* (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* 84(3):428-445.
3. Blokesch M (2012) TransFLP--a method to genetically modify *Vibrio cholerae* based on natural transformation and FLP-recombination. *J Vis Exp* (68).
4. Skorupski K & Taylor RK (1996) Positive selection vectors for allelic exchange. *Gene* 169(1):47-52.
5. Papenfort K, *et al.* (2017) A *Vibrio cholerae* autoinducer-receptor pair that controls biofilm formation. *Nat Chem Biol* 13(5):551-557.
6. Baba T, *et al.* (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2:2006 0008.
7. Papenfort K, *et al.* (2006) SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol* 62(6):1674-1688.
8. Frohlich KS, Haneke K, Papenfort K, & Vogel J (2016) The target spectrum of SdsR small RNA in *Salmonella*. *Nucleic Acids Res* 44(21):10406-10422.
9. Papenfort K, Forstner KU, Cong JP, Sharma CM, & Bassler BL (2015) Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proc Natl Acad Sci U S A* 112(7):E766-775.
10. Frohlich KS, Papenfort K, Fekete A, & Vogel J (2013) A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J* 32(22):2963-2979.
11. Thelin KH & Taylor RK (1996) Toxin-coregulated pilus, but not mannose-sensitive hemagglutinin, is required for colonization by *Vibrio cholerae* O1 El Tor biotype and O139 strains. *Infect Immun* 64(7):2853-2856.
12. Svenningsen SL, Tu KC, & Bassler BL (2009) Gene dosage compensation calibrates four regulatory RNAs to control *Vibrio cholerae* quorum sensing. *EMBO J* 28(4):429-439.
13. Peschek N, Hoyos M, Herzog R, Forstner KU, & Papenfort K (2019) A conserved RNA seed-pairing domain directs small RNA-mediated stress resistance in enterobacteria. *EMBO J* 38(16):e101650.
14. de Lorenzo V & Timmis KN (1994) Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons. *Methods Enzymol* 235:386-405.
15. Herzog R, Peschek N, Frohlich KS, Schumacher K, & Papenfort K (2019) Three autoinducer molecules act in concert to control virulence gene expression in *Vibrio cholerae*. *Nucleic Acids Res* 47(6):3171-3183.



## 3 Gene autoregulation by 3' UTR-derived bacterial small RNAs

Mona Hoyos, **Michaela Huber**, Konrad U. Förstner, and Kai Papenfort (2020).  
“Gene autoregulation by 3' UTR-derived bacterial small RNAs.” *eLife* 9, e58836.

[doi.org/10.7554/eLife.58836](https://doi.org/10.7554/eLife.58836)



# Gene autoregulation by 3' UTR-derived bacterial small RNAs

Mona Hoyos<sup>1,2</sup>, Michaela Huber<sup>1,2</sup>, Konrad U Förstner<sup>3,4</sup>, Kai Papenfort<sup>1,2,5\*</sup>

<sup>1</sup>Friedrich Schiller University Jena, Institute of Microbiology, Jena, Germany;

<sup>2</sup>Faculty of Biology I, Ludwig-Maximilians-University of Munich, Martinsried, Germany;

<sup>3</sup>TH Köln - University of Applied Sciences, Institute of Information Science, Cologne, Germany;

<sup>4</sup>ZB MED - Information Centre for Life Sciences, Cologne, Germany;

<sup>5</sup>Microverse Cluster, Friedrich Schiller University Jena, Jena, Germany

**Abstract** Negative feedback regulation, that is the ability of a gene to repress its own synthesis, is the most abundant regulatory motif known to biology. Frequently reported for transcriptional regulators, negative feedback control relies on binding of a transcription factor to its own promoter. Here, we report a novel mechanism for gene autoregulation in bacteria relying on small regulatory RNA (sRNA) and the major endoribonuclease, RNase E. TIER-seq analysis (transiently-inactivating-an-endoribonuclease-followed-by-RNA-seq) revealed ~25,000 RNase E-dependent cleavage sites in *Vibrio cholerae*, several of which resulted in the accumulation of stable sRNAs. Focusing on two examples, OppZ and CarZ, we discovered that these sRNAs are processed from the 3' untranslated region (3' UTR) of the *oppABCDF* and *carAB* operons, respectively, and base-pair with their own transcripts to inhibit translation. For OppZ, this process also triggers Rho-dependent transcription termination. Our data show that sRNAs from 3' UTRs serve as autoregulatory elements allowing negative feedback control at the post-transcriptional level.

\*For correspondence:

kai.papenfort@uni-jena.de

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 23

**Received:** 12 May 2020

**Accepted:** 23 July 2020

**Published:** 03 August 2020

**Reviewing editor:** Joseph T Wade, Wadsworth Center, New York State Department of Health, United States

© Copyright Hoyos et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

Biological systems function on a mechanism of inputs and outputs, each triggered by and triggering a specific response. Feedback control (a.k.a. autoregulation) is a regulatory principle wherein the output of a system amplifies (positive feedback) or reduces (negative feedback) its own production. Negative feedback regulation is ubiquitous among biological systems and belongs to the most thoroughly characterized network motifs (*Nitzan et al., 2017; Shen-Orr et al., 2002*). At the gene regulatory level, negative feedback control has been qualitatively and quantitatively studied. Most commonly, a transcription factor acts to repress its own transcription by blocking access of RNA polymerase to the promoter region. This canonical mode of negative autoregulation is universally present in living systems and in *Escherichia coli* more than 40% of the known transcription factors are controlled by this type of regulation (*Rosenfeld et al., 2002*). Several characteristics have been attributed to negative autoregulatory circuits including an altered response time and improved robustness towards fluctuations in transcript production rates (*Alon, 2007*).

More recently, the mechanisms underlying RNA-based gene regulation have also been investigated for their regulatory principles and network functions (*Nitzan et al., 2017; Pu et al., 2019*). In bacteria, small regulatory RNAs (sRNAs) constitute the largest class of RNA regulators and frequently bind to one of the major RNA-binding proteins, Hfq or ProQ. Hfq- and ProQ-associated sRNAs usually act by base-pairing with *trans*-encoded target mRNAs affecting translation initiation and transcript stability (*Holmqvist and Vogel, 2018; Kavita et al., 2018*). The sRNAs frequently target multiple transcripts and given that regulation can involve target repression or activation, it has

become ever more clear that sRNAs can rival transcription factors with respect to their regulatory scope and function (Hör *et al.*, 2018).

Another key factor involved in post-transcriptional gene regulation is ribonuclease E (RNase E), an essential enzyme in *E. coli* and related bacteria required for ribosome biogenesis and tRNA maturation (Mackie, 2013). RNase E's role in sRNA-mediated expression control is manifold and includes the processing of sRNAs into functional regulators (Chao *et al.*, 2017; Dar and Sorek, 2018a; Papenfort *et al.*, 2015a; Updegrove *et al.*, 2019; Chao *et al.*, 2012) as well as the degradation of target transcripts (Massé *et al.*, 2003; Morita *et al.*, 2005). Inhibition of RNase E-mediated cleavage through sRNAs can stabilize the target transcript and activate gene expression (Fröhlich *et al.*, 2013; Papenfort *et al.*, 2013; Richards and Belasco, 2019).

Global transcriptome analyses have revealed the presence of numerous sRNAs produced from 3' UTRs (untranslated regions) of mRNAs, a significant subset of which requires RNase E for their maturation (Adams and Storz, 2020). These 3' UTR-derived sRNAs can be produced from monocistronic (Chao and Vogel, 2016; Grabowicz *et al.*, 2016; Huber *et al.*, 2020; Wang *et al.*, 2020) as well as long, operonic mRNAs (Davis and Waldor, 2007; De Mets *et al.*, 2019; Miyakoshi *et al.*, 2019) and typically act to regulate multiple target mRNAs in trans. The RNase E C-terminus also provides the scaffold for a large protein complex, called the degradosome, which in the major human pathogen, *Vibrio cholerae*, has recently been implicated in the turn-over of hypomodified tRNA species (Kimura and Waldor, 2019).

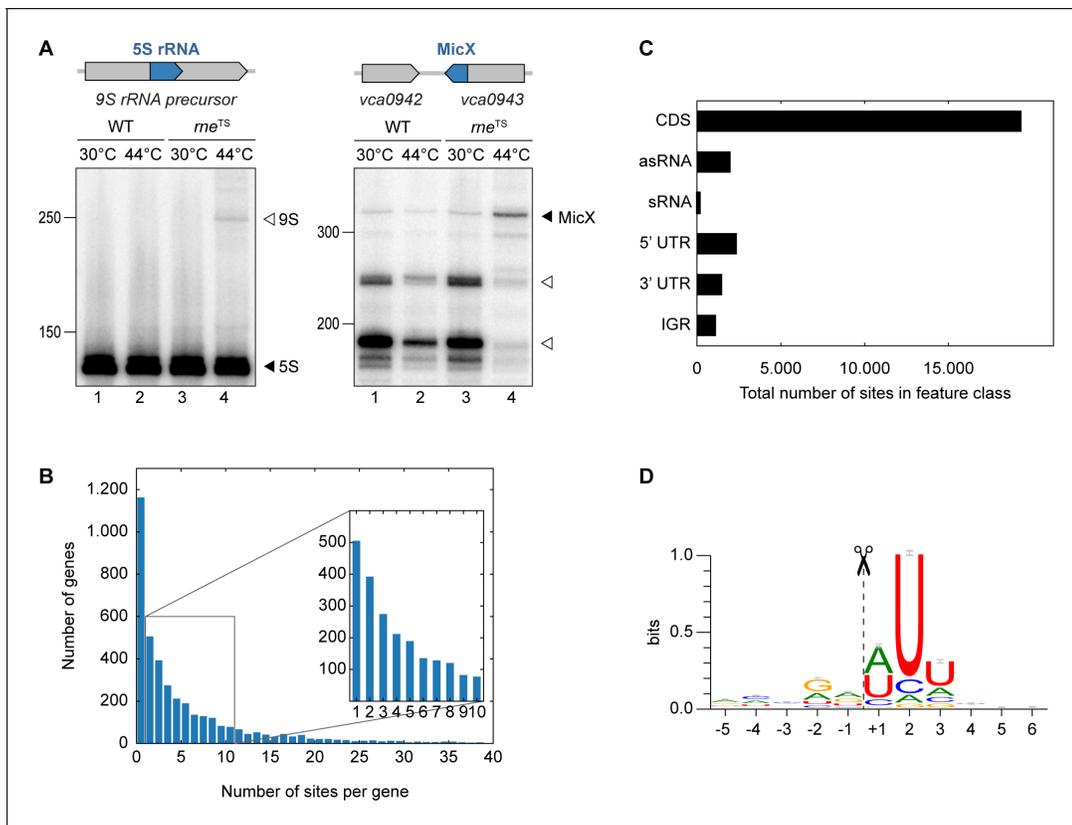
The present work addresses the regulatory role of RNase E in *V. cholerae* at a genome-wide level. To this end, we generated a temperature-sensitive variant of RNase E in *V. cholerae* and employed TIER-seq (transiently-inactivating-an-endoribonuclease-followed-by-RNA-seq) to globally map RNase E cleavage sites (Chao *et al.*, 2017). Our analyses identified ~25,000 RNase E-sensitive sites and revealed the presence of numerous stable sRNAs originating from the 3' UTR of coding sequences. Detailed analyses of two of these sRNAs, OppZ and CarZ, showed that 3' UTR-derived sRNAs can act in an autoregulatory manner to reduce the expression of mRNAs produced from the same genetic locus. The molecular mechanism of sRNA-mediated gene autoregulation likely involves inhibition of translation initiation by the sRNA followed by Rho-dependent transcription termination. This setup directly links the regulatory activity of the sRNAs to their de novo synthesis, analogous to their transcription factor counterparts. However, we show that, in contrast to transcriptional regulators, autoregulatory RNAs can act at a subcistronic level to allow discoordinate operon expression.

## Results

### TIER-seq analysis of *V. cholerae*

The catalytic activity of RNase E (encoded by the *rne* gene) is critical for many bacteria, including *V. cholerae* (Cameron *et al.*, 2008). To study the role of RNase E in this pathogen, we mutated the DNA sequence of the *V. cholerae* chromosome encoding leucine 68 of RNase E to phenylalanine (Figure 1—figure supplement 1). This mutation is analogous to the originally described N3071 *rne*<sup>TS</sup> isolate of *E. coli* (Apirion and Lassar, 1978) and exhibits full RNase E activity at permissive temperatures (30°C), but is rendered inactive under non-permissive temperatures (44°C). We validated our approach by monitoring the expression of two known substrates of RNase E in *V. cholerae*: A) 5S rRNA, which is processed by RNase E from the 9S precursor rRNA (Papenfort *et al.*, 2015b), and B) the MicX sRNA, which contains two RNase E cleavage sites (Davis and Waldor, 2007). For both RNAs, transfer of the wild-type strain to 44°C only mildly effected their expression, whereas the equivalent procedure performed with the *rne*<sup>TS</sup> strain led to the accumulation of the 9S precursor and the full-length MicX transcript (Figure 1A, lanes 1–2 vs. 3–4). Additionally, accumulation of the two RNase E-dependent processing intermediates of MicX was reduced in the *rne*<sup>TS</sup> strain at the non-permissive temperature.

These results showed that we successfully generated a temperature-sensitive RNase E variant in *V. cholerae* and enabled us to employ TIER-seq to determine RNase E-dependent cleavage sites at a global scale. To this end, we cultivated *V. cholerae* wild-type and *rne*<sup>TS</sup> strains at 30°C to late exponential phase (OD<sub>600</sub> of 1.0), divided the cultures in half and continued incubation for 60 min at either 30°C or 44°C. Total RNA was isolated and subjected to deep sequencing. We obtained ~187 million reads from the twelve samples (corresponding to three biological replicates of each strain



**Figure 1.** TIER-seq analysis of *V. cholerae*. (A) *V. cholerae* wild-type and *rne<sup>TS</sup>* strains were grown at 30°C to stationary phase (OD<sub>600</sub> of 2.0). Cultures were divided in half and continuously grown at either 30°C or 44°C for 60 min. Cleavage patterns of 5S rRNA and 3' UTR-derived MicX were analyzed on Northern blots. Closed triangles indicate mature 5S or full-length MicX, open triangles indicate the 9S precursor or MicX processing products. (B, C, D) Biological triplicates of *V. cholerae* wild-type and *rne<sup>TS</sup>* strains were grown at 30°C to late exponential phase (OD<sub>600</sub> of 1.0). Cultures were divided in half and continuously grown at either 30°C or 44°C for 60 min. Isolated RNA was subjected to RNA-seq and RNase E cleavage sites were determined as described in the materials and methods section. (B) Number of cleavage sites detected per gene. (C) Classification of RNase E sites by their genomic location. (D) The RNase E consensus motif based on all detected cleavage sites. The total height of the error bar is twice the small sample correction. The online version of this article includes the following source data and figure supplement(s) for figure 1:

**Source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 1** and RNase E cleavage site counts within genes or transcript categories.

**Figure supplement 1.** Conservation of RNase E between *E. coli* and *V. cholerae*.

**Figure supplement 2.** TIER-Seq read mapping statistics.

**Figure supplement 2—source data 1.** Number of obtained sequencing reads and Pearson correlation coefficients for library comparisons.

**Figure supplement 3.** Position and characteristics of RNase E cleavage sites.

**Figure supplement 4.** RNase E-mediated maturation of sRNAs from 3' UTRs.

**Figure supplement 4—source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 1—figure supplement 4**.

**Figure supplement 5.** RNase E-mediated maturation of sRNAs from IGRs.

**Figure supplement 5—source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 1—figure supplement 5**.

**Figure supplement 6.** Expression of RNase E-independent sRNAs.

**Figure supplement 6—source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 1—figure supplement 6**.

and condition; **Figure 1—figure supplement 2A**), resulting in ~98 million unique 5' ends mapping to the *V. cholerae* genome. Comparison of the 5' ends detected in wild-type and *rne<sup>TS</sup>* at 30°C showed almost no difference between the two strains (Pearson correlation coefficients  $R^2$  ranging from 0.82 to 0.99 depending on the compared replicates), whereas the same analysis at 44°C revealed 24,962 depleted sites in the *rne<sup>TS</sup>* strain (**Figure 1—figure supplement 2B–C**). Given that  $\gamma$ -proteobacteria such as *V. cholerae* do not encode 5' to 3' exoribonucleases (**Mohanty and Kushner, 2018**), we designated these positions as RNase E-specific cleavage sites (**Supplementary file 1**).

Next, we analysed the ~25,000 RNase E sites with respect to frequency per gene and their distribution among different classes of transcript. We discovered that RNase E cleavage sites occur with a frequency of 2.8 (median)/6.3 (mean) sites per kb (**Figure 1B**). The majority of cleavage events occurs in coding sequences (~69.1%), followed by 5' UTRs (~8.4%), antisense RNAs (~7.1%), 3' UTRs (~5.3%), intergenic regions (~4.0%), and sRNAs (~0.6%) (**Figure 1C**). RNase E sites were slightly enriched around start and stop codons of mRNAs (**Figure 1—figure supplement 3A**). Furthermore, cleavage coincided with an increase in AU-content (**Figure 1—figure supplement 3B**) and a rise in minimal folding energies (**Figure 1—figure supplement 3C**), suggesting reduced secondary structure around RNase E sites. Together, these data allowed us to determine a consensus motif for RNase E in *V. cholerae* (**Figure 1D**). This 5-nt sequence, i.e. 'RN↓WUU', is highly similar to previously determined RNase E motifs of *Salmonella enterica* (**Chao et al., 2017**) and *Rhodobacter sphaeroides* (**Förstner et al., 2018**), indicating that RNase E operates by a conserved mechanism of recognition and cleavage.

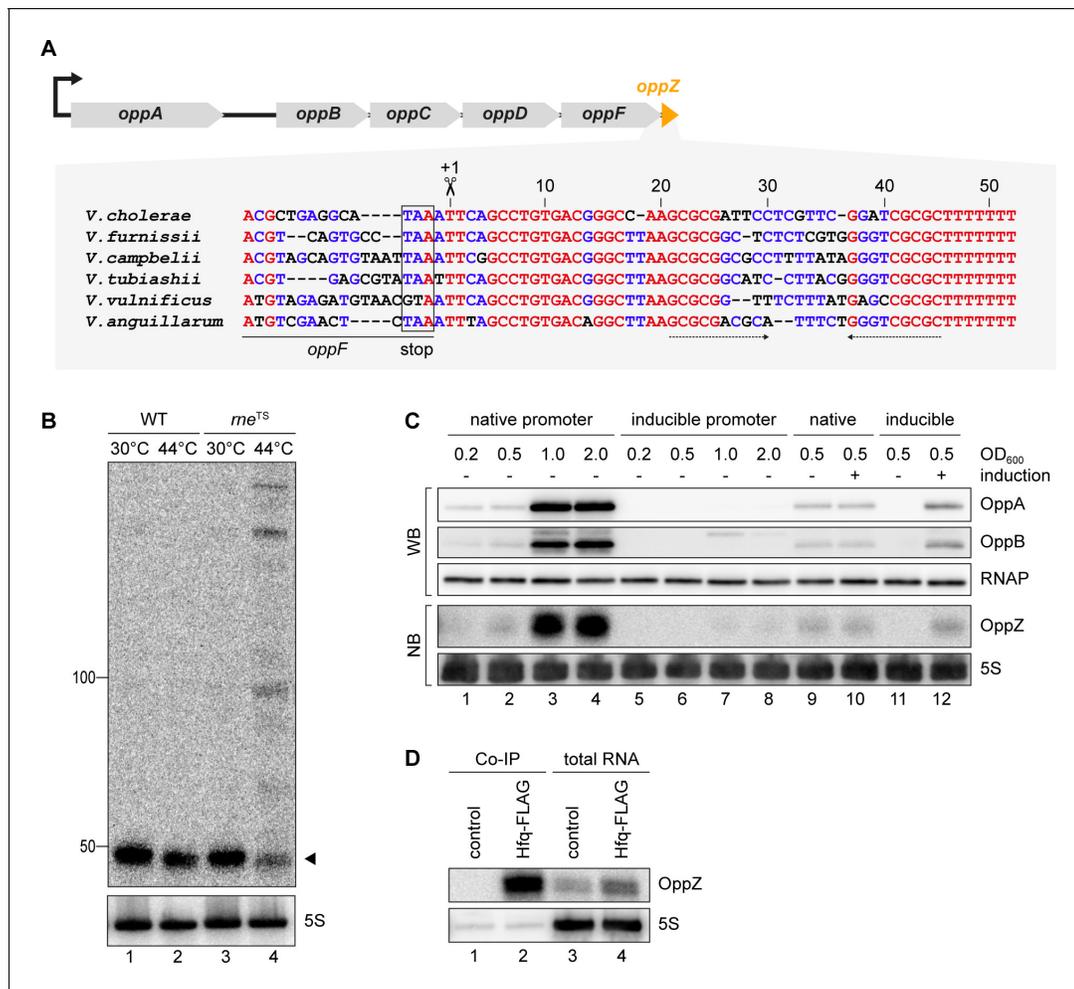
## RNase E-mediated maturation of sRNAs

Earlier work on sRNA biogenesis in bacteria revealed that the 3' UTR of coding transcripts can serve as source for non-coding regulators and that RNase E is frequently required to cleave the sRNA from the mRNA (**Miyakoshi et al., 2015**). In *V. cholerae*, we previously annotated 44 candidate sRNAs located in the 3' UTR of mRNAs (**Papenfort et al., 2015b**). To analyse which of these sRNAs depend on RNase E for maturation, we searched for RNase E-cleavage sites matching with the first three bases of the annotated sRNAs. 17 sRNAs revealed potential RNase E-dependent maturation (**Supplementary file 2A**) and using Northern blot analyses of wild-type and *rne*<sup>TS</sup> samples, we were able to confirm these results for 9 sRNAs (Vcr016, Vcr041, Vcr044, Vcr045, Vcr053, Vcr064, FarS, Vcr079, and Vcr084; **Figure 1—figure supplement 4**). In all cases, transfer of the *rne*<sup>TS</sup> strain to non-permissive temperatures led to a change in mature sRNA levels and/or their upstream processing intermediates. We also discovered several sRNAs undergoing maturation by RNase E (**Supplementary file 2B**). Specifically, Northern blot analysis of Vcr043, Vcr065, and Vcr082 revealed that these sRNAs accumulate as multiple stable intermediates (**Figure 1—figure supplement 5**) that may contain different regulatory capacities as previously described for ArcZ and RprA of *S. enterica* (**Chao et al., 2017; Papenfort et al., 2015a; Soper et al., 2010**). In addition, we also analysed the expression of several RNase E-independent sRNAs (RyhB, Spot 42 and VqmR; **Figure 1—figure supplement 6**) on Northern blots. Inactivation of RNase E did not affect the levels of the mature sRNAs or any processed intermediates.

## OppZ is produced from the oppABCDF 3' end

To understand the regulatory functions of 3' UTR-derived sRNAs in *V. cholerae*, we focussed on Vcr045, which is processed from the 3' end of the *oppABCDF* mRNA (encoding an oligopeptide transporter) and which we hence named OppZ. The *oppZ* gene is 52 bps long and conserved among the *Vibrios* (**Figure 2A**). RNase E-mediated cleavage of *oppABCDF* occurs immediately downstream of the *oppF* stop codon and using the *rne*<sup>TS</sup> strain, we were able to validate RNase E-dependent processing of OppZ (**Figure 2B**). Northern and Western blot analysis of a *V. cholerae* strain carrying a 3XFLAG epitope at the C-terminus of the chromosomal *oppA* and *oppB* genes revealed that OppZ expression coincided with the expression of both proteins (**Figure 2C**, lanes 1–4). Previous transcriptome data showed that expression of *oppABCDF* is controlled by a single promoter located ~120 bps upstream of *oppA* (**Papenfort et al., 2015b**), indicating that the sRNA is co-expressed with all five *opp* genes. To test this prediction, we replaced the native promoter upstream of the chromosomal *oppA* gene with the L-arabinose-inducible pBAD promoter and monitored OppA, OppB, and OppZ expression under inducing and non-inducing conditions. In the absence of the inducer, expression of OppA/B and OppZ was strongly reduced (**Figure 2C**, lanes 5–8) and L-arabinose had no effect on the activity of the native *oppA* promoter (**Figure 2C**, lanes 9–10). In contrast, activation of the pBAD promoter led to a significant increase in OppA/B and OppZ (**Figure 2C**, lanes 11–12), indicating that expression of the *oppABCDF-oppZ* operon is indeed controlled by a single promoter.

To support these results and confirm production of OppZ from the longer precursor transcript, we generated two plasmids carrying either only *oppZ* or *oppF-oppZ* under the control of the



**Figure 2.** OppZ is produced from the *oppABCDF* 3' end. (A) Top: Genomic organization of *oppABCDF* and *oppZ*. Bottom: Alignment of *oppZ* sequences, including the last codons of *oppF*, from various *Vibrio* species. The *oppF* stop codon, the RNase E cleavage site and the Rho-independent terminator are indicated. (B) *V. cholerae* wild-type and *rne<sup>TS</sup>* strains were grown at 30°C to stationary phase (OD<sub>600</sub> of 2.0). Cultures were divided in half and continuously grown at either 30°C or 44°C for 30 min. OppZ synthesis was analyzed by Northern blot with 5S rRNA as loading control. The triangle indicates the size of mature OppZ. (C) Protein and RNA samples were obtained from *V. cholerae* *oppA::3XFLAG oppB::3XFLAG* strains carrying either the native *oppA* promoter or the inducible pBAD promoter upstream of *oppA*. Samples were collected at the indicated OD<sub>600</sub> and tested for OppA and OppB production by Western blot and for OppZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. Lanes 1–8: Growth without L-arabinose. Lanes 9–12: Growth with either H<sub>2</sub>O (-) or L-arabinose (+) (0.2% final conc.). (D) *V. cholerae* wild-type (control) and *hfq::3XFLAG* (Hfq-FLAG) strains were grown to stationary phase (OD<sub>600</sub> of 2.0), lysed, and subjected to immunoprecipitation using the anti-FLAG antibody. RNA samples of lysate (total RNA) and co-immunoprecipitated fractions were analyzed on Northern blots. 5S rRNA served as loading control.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

**Source data 1.** Full Northern and Western blot images for the corresponding detail sections shown in **Figure 2**.

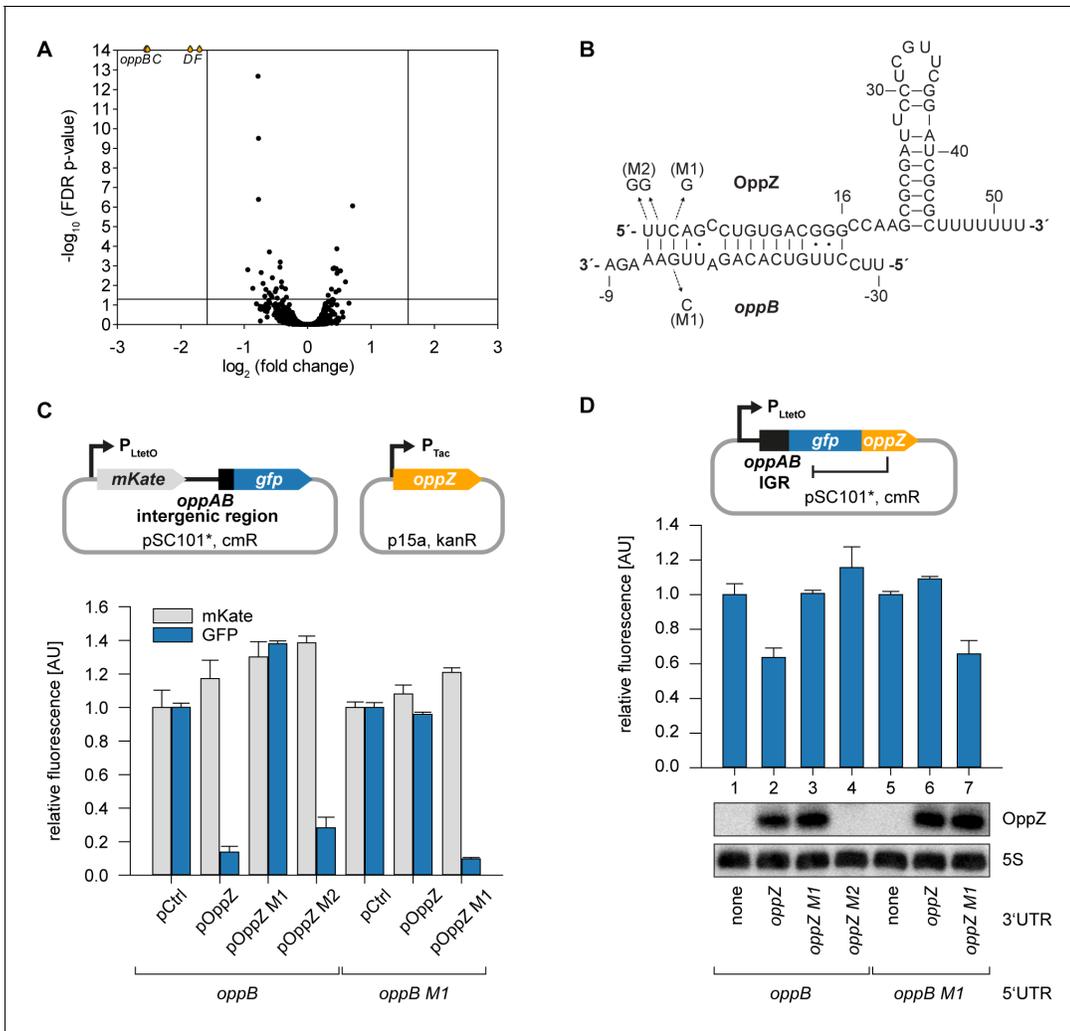
**Figure supplement 1.** Hfq dependence of OppZ processing.

**Figure supplement 1—source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 2—figure supplement 1**.

**Figure supplement 2.** Hfq dependence of OppZ stability.

**Figure supplement 2—source data 1.** Quantification of OppZ levels in wild-type and  $\Delta$ hfq cells from Northern blots.

constitutive P<sub>Tac</sub> promoter (**Figure 2—figure supplement 1A**) and compared OppZ expression in wild-type and  $\Delta$ oppZ cells. Expression of mature OppZ was readily detected from the precursor (**Figure 2—figure supplement 1B**, lane 1 vs. 4) and the size of the processed OppZ transcript was comparable to endogenously expressed OppZ (lane 1) and OppZ transcribed directly by the P<sub>Tac</sub> promoter (lane 3). We also repeated these experiments in a *V. cholerae* *hfq* mutant (Svenningsen et al., 2009). Here, processing of the precursor into OppZ was still detected



**Figure 3.** Feedback autoregulation at the suboperonic level. (A) Volcano plot of genome-wide transcript changes in response to inducible OppZ over-expression. Lines indicate cut-offs for differentially regulated genes at 3-fold regulation and FDR-adjusted p-value  $\leq 0.05$ . Genes with an FDR-adjusted p-value  $< 10^{-14}$  are indicated as droplets at the top border of the graph. (B) Predicted OppZ secondary structure and base-pairing to *oppB*. Arrows indicate the mutations tested in (C) and (D). (C) *E. coli* strains carrying a translational reporter plasmid with the *oppAB* intergenic region placed between *mKate2* and *gfp* were co-transformed with a control plasmid or the indicated OppZ expression plasmids. Transcription of the reporter and *oppZ* were driven by constitutive promoters. Cells were grown to  $OD_{600} = 1.0$  and fluorophore production was measured. *mKate* and GFP levels of strains carrying the control plasmid were set to 1. Error bars represent the SD of three biological replicates. (D) Single-plasmid regulation was measured by inserting the indicated *oppZ* variant into the 3' UTR of a translational *oppB::gfp* fusion. Expression was driven from a constitutive promoter. *E. coli* strains carrying the respective plasmids were grown to  $OD_{600} = 1.0$  and GFP production was measured. Fluorophore levels from control fusions without an sRNA gene were set to one and error bars represent the SD of three biological replicates. OppZ expression was tested by Northern blot; 5S rRNA served as loading control.

The online version of this article includes the following source data and figure supplement(s) for figure 3:

**Source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 3** and raw data for fluorescence measurements.

**Figure supplement 1.** Pulse expression of OppZ reduces *oppBCDF* transcript levels.

**Figure supplement 1—source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 3—figure supplement 1** and raw data for transcript changes as determined by qRT-PCR.

**Figure supplement 2.** Hfq-dependent, post-transcriptional repression of OppBCDF by OppZ.

**Figure supplement 2—source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 3—figure supplement 2** and raw data for fluorescence measurements.

**Figure supplement 3.** Mutational analysis of the RNase E site in *oppZ*.

**Figure supplement 3—source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 3—figure supplement 3**.

(lane 8), however, the steady-state levels of OppZ were lower, suggesting that OppZ binds Hfq. Indeed, stability experiments using rifampicin-treated *V. cholerae* showed that OppZ half-life is reduced in  $\Delta hfq$  cells (**Figure 2—figure supplement 2**), and RNA co-immunoprecipitation experiments of chromosomal Hfq::3XFLAG revealed that OppZ interacts with Hfq in vivo (**Figure 2D**). Together, these data show that OppZ is an Hfq-dependent sRNA that is processed from the 3' UTR of the polycistronic *oppABCDF* mRNA by RNase E.

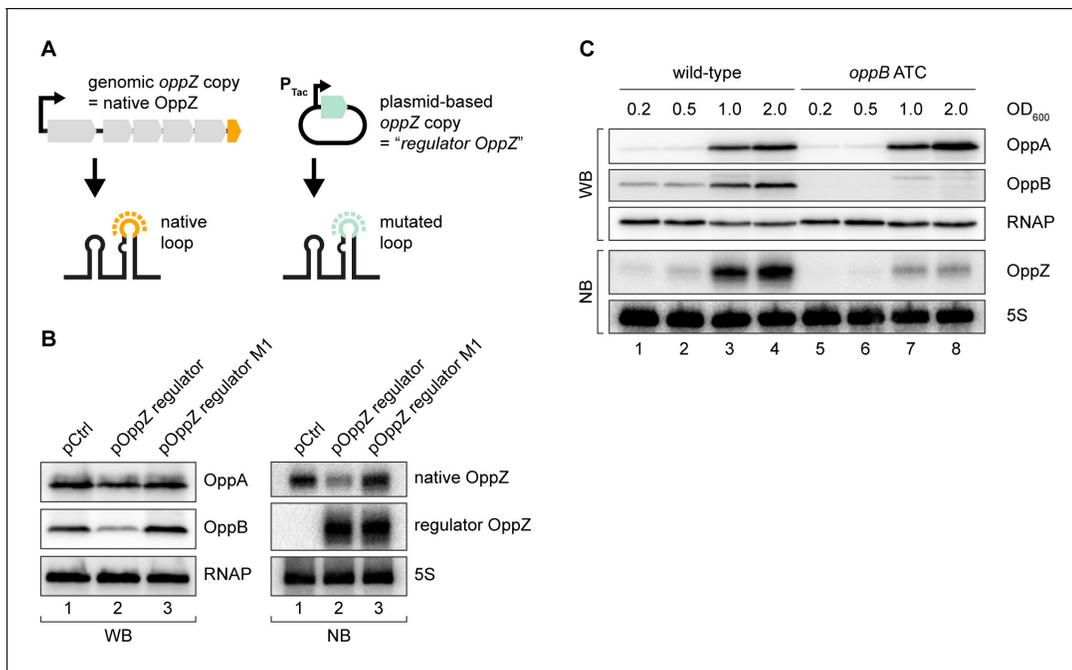
## Feedback Autoregulation at the suboperonic level

Hfq-binding sRNAs control gene expression by base-pairing with *trans*-encoded target transcripts (**Kavita et al., 2018**). To determine the targets of OppZ in *V. cholerae*, we cloned the sRNA (starting from the RNase E cleavage site) on a plasmid under the control of the pBAD promoter. Induction of the pBAD promoter for 15 min resulted in a strong increase in OppZ levels (~30 fold, **Figure 3—figure supplement 1A**) and RNA-seq experiments of the corresponding samples revealed four repressed genes (**Figure 3A** and **Figure 3—figure supplement 1B**). Interestingly, these genes were *oppBCDF*, i.e. the same transcript that OppZ is processed from. We validated OppZ-mediated repression of all four genes using qRT-PCR (**Figure 3—figure supplement 1C**), which also confirmed that the first gene of the operon, *oppA*, is not affected by OppZ. Despite the reduced transcript levels of *oppBCDF*, OppZ over-expression did not reduce the stability of the *oppB* messenger (**Figure 3—figure supplement 1D**). Using the RNA-hybrid algorithm (**Rehmsmeier et al., 2004**), we were able to predict RNA duplex formation of the *oppB* translation initiation site with the 5' end of the OppZ sRNA (**Figure 3B**). We confirmed this interaction using a variant of a previously reported post-transcriptional reporter system (**Corcoran et al., 2012**). Here, the first gene of the operon is replaced by the red-fluorescent mKate2 protein, followed by the *oppAB* intergenic sequence and the first five codons of *oppB*, which were fused to *gfp* (**Figure 3C**, top). Transfer of this plasmid into *E. coli* and co-transformation of the OppZ over-expression plasmid resulted in strong repression of GFP (~7 fold), while mKate2 levels remained constant. Mutation of either OppZ or *oppB* (mutations M1, see **Figure 3B**) abrogated regulation of GFP and combination of both mutants restored control (**Figure 3C**, bottom). In contrast, OppZ-mediated repression of OppB::GFP was strongly reduced in *E. coli* lacking *hfq* (**Figure 3—figure supplement 2A–B**). We also generated three additional variants of the reporter plasmids in which we included the *oppBC*, *oppBCD*, and *oppBCDF* sequences fused to GFP (**Figure 3—figure supplement 2C**). In all cases, OppZ readily inhibited GFP but did not affect mKate2. These results confirm that OppZ promotes discoordinate expression of the *oppABCDF* operon.

Next, we aimed to reproduce OppZ-mediated repression from a single transcript. To this end, we compared GFP production of a translational *oppB::gfp* reporter with the same construct carrying the *oppZ* sequence downstream of *gfp* (**Figure 3D**, top). Northern blot analysis revealed that OppZ was efficiently clipped off from the *gfp* transcript in this construct and fluorescence measurements showed that OppZ also inhibited GFP expression (**Figure 3D**, bottom, lane 1 vs. 2). We confirmed that this effect is specific to base-pairing of OppZ with the *oppAB* intergenic sequence as we were able to recapitulate our previous compensatory base-pair exchange experiments using the single plasmid system (**Figure 3D**). In addition, mutation of the RNase E recognition site in *oppZ* (UU→GG, mutation M2; **Figure 3—figure supplement 3A**) blocked OppZ maturation and abolished OppB::GFP repression (**Figure 3D**, lane 4; **Figure 3—figure supplement 3B**), whereas expression of OppZ M2 from a separate plasmid efficiently reduced OppB:GFP levels (**Figure 3C**). Together, our data demonstrate that OppZ down-regulates protein synthesis from its own cistron. Furthermore, mutation M2 shows that this autoregulation is not mediated by long-distance intramolecular base-pairing of OppZ with the *oppB* 5' UTR, but rather requires RNase E-dependent maturation of the transcript followed by Hfq-dependent base-pairing.

## Translational control of OppZ synthesis

The above experiments revealed that OppZ inhibits protein production through feedback control, however, it was not clear if OppZ would also inhibit its own synthesis. To address this question, we generated an OppZ over-expression plasmid in which we mutated the sequence of the terminal stem-loop at eight positions. We call this construct 'regulator OppZ' (**Figure 4A**). These mutations are not expected to inactivate the base-pairing function of OppZ, but will allow us to differentiate



**Figure 4.** Translational control of OppZ synthesis. (A) Schematic of the analyzed OppZ variants containing the native stem loop sequence (produced from the genomic *oppZ* locus) or a mutated stem loop sequence (*regulator OppZ* produced from a plasmid-based constitutive promoter). (B) *V. cholerae oppA::3XFLAG oppB::3XFLAG* carrying a control plasmid (pCMW-1) or a plasmid expressing *regulator OppZ* (pMD194, pMD195) were grown to stationary phase ( $OD_{600}$  of 2.0). OppA and OppB production were tested by Western blot and expression of native OppZ and regulator OppZ was monitored on Northern blot using oligonucleotides binding to the respective loop sequence variants. RNAP and 5S rRNA served as loading controls for Western blot and Northern blot, respectively. (C) The *oppB* start codon was mutated to ATC in an *oppA::3XFLAG oppB::3XFLAG* background. *V. cholerae* strains with wild-type or mutated *oppB* start codon were grown in LB medium. Protein and RNA samples were collected at the indicated  $OD_{600}$  and tested for OppA and OppB production by Western blot and for OppZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively.

The online version of this article includes the following source data and figure supplement(s) for figure 4:

**Source data 1.** Full Northern and Western blot images for the corresponding detail sections shown in **Figure 4**.

**Figure supplement 1.** Translational control of OppZ synthesis.

**Figure supplement 1—source data 1.** Quantification of OppZ levels in wild-type and *oppB* ATC cells from Northern blots and full blot images for the corresponding detail sections shown in **Figure 4—figure supplement 1**.

the levels of native OppZ and *regulator OppZ* on Northern blots. Indeed, when tested in *V. cholerae*, over-expression of *regulator OppZ* inhibited OppB::3XFLAG production, but did not affect OppA::3XFLAG levels (**Figure 4B**, left). Importantly, *regulator OppZ* also reduced the expression of native OppZ (**Figure 4B**, right) and introduction of the M1 mutation (see **Figure 3B**) in *regulator OppZ* abrogated this effect. These results revealed that OppZ also exerts autoregulation of its own transcript.

Gene expression control by sRNAs typically occurs post-transcriptionally (**Gorski et al., 2017**) raising the question of how OppZ achieves autoregulation at the molecular level. Given that OppZ inhibits OppB production (**Figure 4B**), we hypothesized that OppZ synthesis might be linked to *oppB* translation. To test this prediction, we inactivated the chromosomal start codon of *oppB* (ATG→ATC) and monitored OppA/B and OppZ expression by Western and Northern blot, respectively. As expected, mutation of the *oppB* start codon had no effect on OppA::3XFLAG levels, but nullified OppB::3XFLAG production (**Figure 4C**, top). Lack of *oppB* translation also resulted in a strong decrease in OppZ levels (**Figure 4C**, bottom), however, did not change OppZ stability (**Figure 4—figure supplement 1A**). In addition, plasmid-based complementation of OppB::3XFLAG in the *oppB* start codon mutant failed to restore OppZ expression (**Figure 4—figure supplement 1B**), showing that OppZ production is independent of the cellular OppB levels. Based on these and the results above, we propose that autorepression of *oppBCDF-oppZ* must occur by a mechanism involving both translation inhibition, as well as transcription termination.

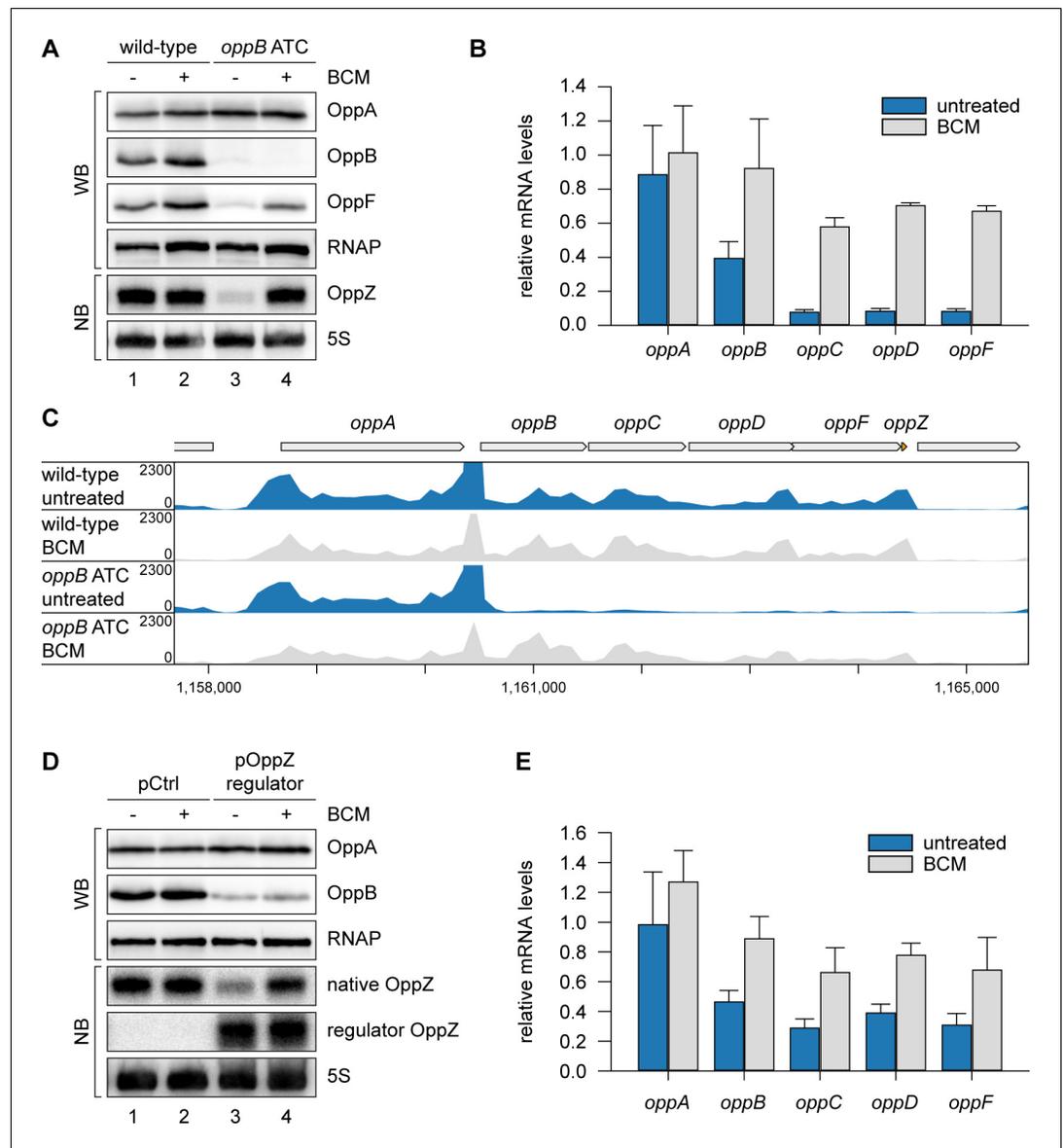
## OppZ promotes transcription termination through Rho

To explain the reduction of OppZ expression in the absence of *oppB* translation, we considered premature transcription termination as a possible factor. This hypothesis was supported by our finding that OppZ over-expression efficiently reduced *oppB* mRNA levels without significantly affecting transcript stability (**Figure 3—figure supplement 1C–D**). In *E. coli*, Rho protein accounts for a major fraction of all transcription termination events (**Ciampi, 2006**) and has previously been associated with the regulatory activity of Hfq-dependent sRNAs (**Bossi et al., 2012; Sedlyarova et al., 2016; Wang et al., 2015**). Rho is specifically inhibited by bicyclomycin (BCM; **Zwiefka et al., 1993**) and consequently we tested the effect of the antibiotic on OppZ expression in *V. cholerae* wild-type and the *oppB* start codon mutant. Whereas BCM had no effect on OppZ synthesis in wild-type cells (**Figure 5A**, lane 1 vs. 2), it strongly increased OppZ and *oppBCDF* expression in the absence of *oppB* translation (**Figure 5A**, lane 3 vs. 4, and **Figure 5B**). We confirmed these results by employing Term-Seq analysis (**Dar et al., 2016**) to wild-type and *oppB* start codon mutants cultivated with or without BCM. Detailed inspection of transcript coverage at the *oppABCDF-oppZ* genomic locus showed that lack of *oppB* translation down-regulated the expression of *oppBCDF-oppZ*, while presence of BCM suppressed this effect (**Figure 5C** and **Supplementary file 3B**). Similarly, inhibition of the *oppBCDF* mRNA and OppZ by over-expression of regulator *OppZ* (see **Figure 4A**) was suppressed in the presence of BCM, whereas *OppB* protein levels remained low presumably due to continued repression of *oppB* translation initiation by OppZ (**Figure 5D–E**).

To map the position of Rho-dependent transcription termination in *oppB*, we generated five additional strains carrying a STOP mutation at the 2<sup>nd</sup>, 15<sup>th</sup>, 65<sup>th</sup>, 115<sup>th</sup>, or 215<sup>th</sup> codon of the chromosomal *oppB* gene (**Figure 6A**). In addition, we mutated the start codons of *oppC*, *oppD*, and *oppF* and probed OppZ levels on Northern blot (**Figure 6B**). In accordance with the data presented in **Figure 4C**, mutation of the *oppB* start codon resulted in strongly decreased OppZ levels (**Figure 6B**, lane 1 vs. 2) and we observed similar results when the STOP mutation was introduced at the 2<sup>nd</sup>, 15<sup>th</sup>, and 65<sup>th</sup> codon of *oppB* (**Figure 6B**, lanes 3–5). In contrast, a STOP mutation at codon 115 led to increased OppZ expression (lane 6) and OppZ levels were fully restored when the STOP was placed at codon 215 of *oppB* (lane 7). Likewise, mutation of the *oppC*, *oppD*, and *oppF* start codons had no effect on OppZ production (**Figure 6B**, lanes 8–10). To summarize, our data indicate that autorepression of the *oppBCDF-oppZ* genes relies on inhibition of *oppB* translation initiation by OppZ, which triggers Rho-dependent transcription termination in the distal part of the *oppB* sequence.

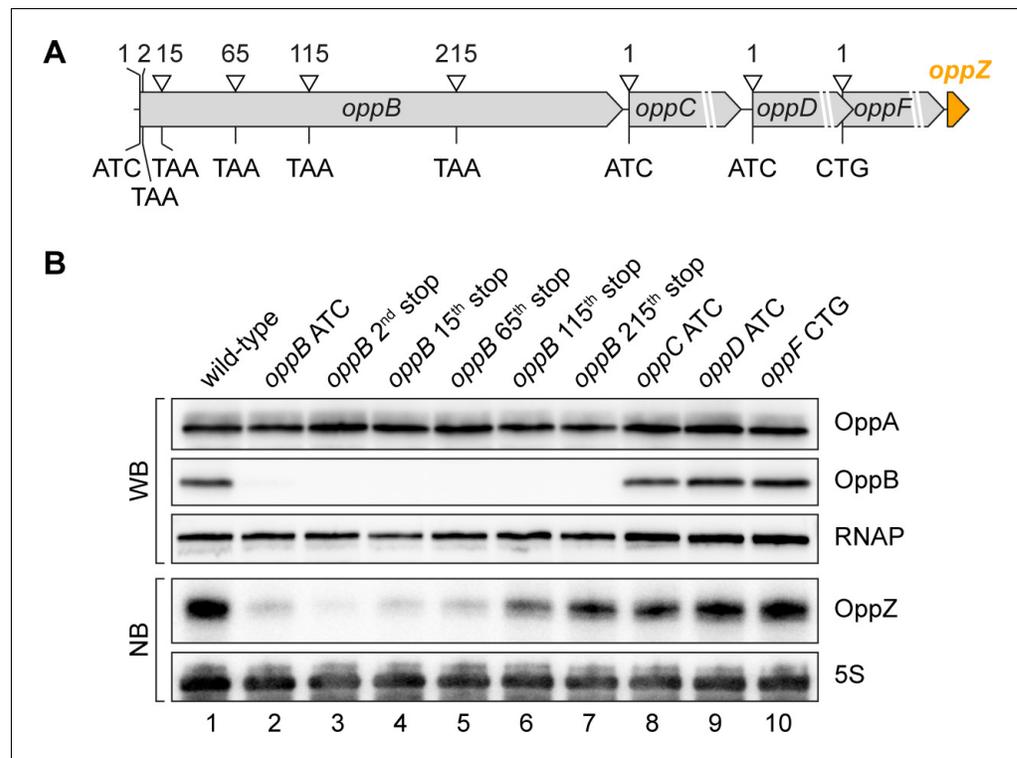
## CarZ is another autoregulatory sRNA from *V. cholerae*

Our TIER-seq analysis revealed 17 3' UTR-derived sRNAs produced by RNase E-mediated cleavage in *V. cholerae* (**Supplementary file 2A**). Detailed analysis of OppZ showed that this sRNA serves as an autoregulatory element inhibiting the *oppBCDF* genes as well as its own synthesis (**Figures 4–6**). We therefore asked how wide-spread RNA-mediated autoregulation is and if the other 16 3' UTR-derived sRNAs might serve a similar function in *V. cholerae*. To this end, we searched for potential base-pairing sequences between the sRNAs and the translation initiation regions of their associated genes using the *RNA-hybrid* algorithm (**Rehmsmeier et al., 2004**). Indeed, we were able to predict stable RNA duplex formation between the Vcr084 sRNA (located in the 3' UTR of the *carAB* operon; encoding carbamoyl phosphate synthetase) and the 5' UTR of *carA*, which is the first gene of the operon (**Figure 7A–B**). In analogy to OppZ, we named this sRNA CarZ. Plasmid-borne expression of CarZ strongly inhibited GFP production from *carA::gfp* and *carAB::gfp* reporters in *E. coli* (**Figure 7—figure supplement 1A–B**) and we obtained similar results using a single transcript *carA::gfp::carZ* construct (**Figure 7C**). CarZ binds Hfq in vivo (**Figure 7—figure supplement 1C**) and repression of *carA::gfp* by CarZ requires Hfq, possibly due to reduced CarZ levels in the *hfq* mutant (**Figure 7—figure supplement 1D–E**). We validated the predicted interaction using compensatory base-pair exchange experiments (**Figure 7B–C**, **Figure 7—figure supplement 1A–B**). Transcription of *carAB-carZ* is controlled by a single promoter located upstream of *carA* and the three genes are co-expressed in vivo (**Figure 7D** and **Papenfort et al., 2015b**). These results suggested that CarZ provides feedback regulation and using an experimental strategy analogous to **Figure 4A**, we were able to show that CarZ inhibits CarA and CarB protein expression as well as its own synthesis (**Figure 7B,E**). Furthermore, introduction of a STOP codon at the 2<sup>nd</sup> codon of the chromosomal



**Figure 5.** OppZ promotes transcription termination through Rho. (A) *V. cholerae* *oppA*::3XFLAG *oppB*::3XFLAG *oppF*::3XFLAG strains with wild-type or mutated *oppB* start codon were grown to early stationary phase (OD<sub>600</sub> of 1.5). Cultures were divided in half and treated with either H<sub>2</sub>O or BCM (25 µg/ml final conc.) for 2 hr before protein and RNA samples were collected. OppA, OppB and OppF production were tested by Western blot and OppZ expression was monitored by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. (B) Biological triplicates of *V. cholerae* *oppA*::3XFLAG *oppB*::3XFLAG strains with wild-type or mutated *oppB* start codon were treated with BCM as described in (A). *oppABCDF* expression in the *oppB* start codon mutant compared to the wild-type control was analyzed by qRT-PCR. Error bars represent the SD of three biological replicates. (C) Triplicate samples from (B) were subjected to Term-seq and average coverage of the *opp* operon is shown for one representative replicate. The coverage cut-off was set at the maximum coverage of annotated genes. (D) *V. cholerae* *oppA*::3XFLAG *oppB*::3XFLAG strains carrying a control plasmid (pMD397) or a plasmid expressing regulator OppZ (pMD398) were treated with BCM as described in (A). OppA and OppB production were tested by Western blot and expression of native OppZ and regulator OppZ was monitored on Northern blot using oligonucleotides binding to the respective loop sequence variants. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. (E) Levels of *oppABCDF* in the experiment described in (D) were analyzed by qRT-PCR. Error bars represent the SD of three biological replicates. The online version of this article includes the following source data for figure 5:

**Source data 1.** Full blot images for the corresponding detail sections shown in **Figure 5** and raw data for transcription changes as determined by qRT-PCR.



**Figure 6.** Influence of OppBCDF translation on OppZ expression. (A) The depicted mutations were individually inserted into the *opp* locus to inactivate the start codons of *oppB*, *oppC*, *oppD* or *oppF* or to insert STOP codons at the positions 2, 15, 65, 115 or 215 of *oppB*. (B) *V. cholerae* *oppA*::3XFLAG *oppB*::3XFLAG strains with the described *opp* mutations were grown: wild-type (lane 1), the *oppB* start codon mutated (lane 2), a STOP codon inserted at the 2<sup>nd</sup>, 15<sup>th</sup>, 65<sup>th</sup>, 115<sup>th</sup> or 215<sup>th</sup> codon of *oppB* (lanes 3–7) or mutated start codons of *oppC*, *oppD* or *oppF* (lanes 8–10). At stationary phase (OD<sub>600</sub> of 2.0), protein and RNA samples were collected and tested for OppA and OppB production by Western blot and for OppZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively.

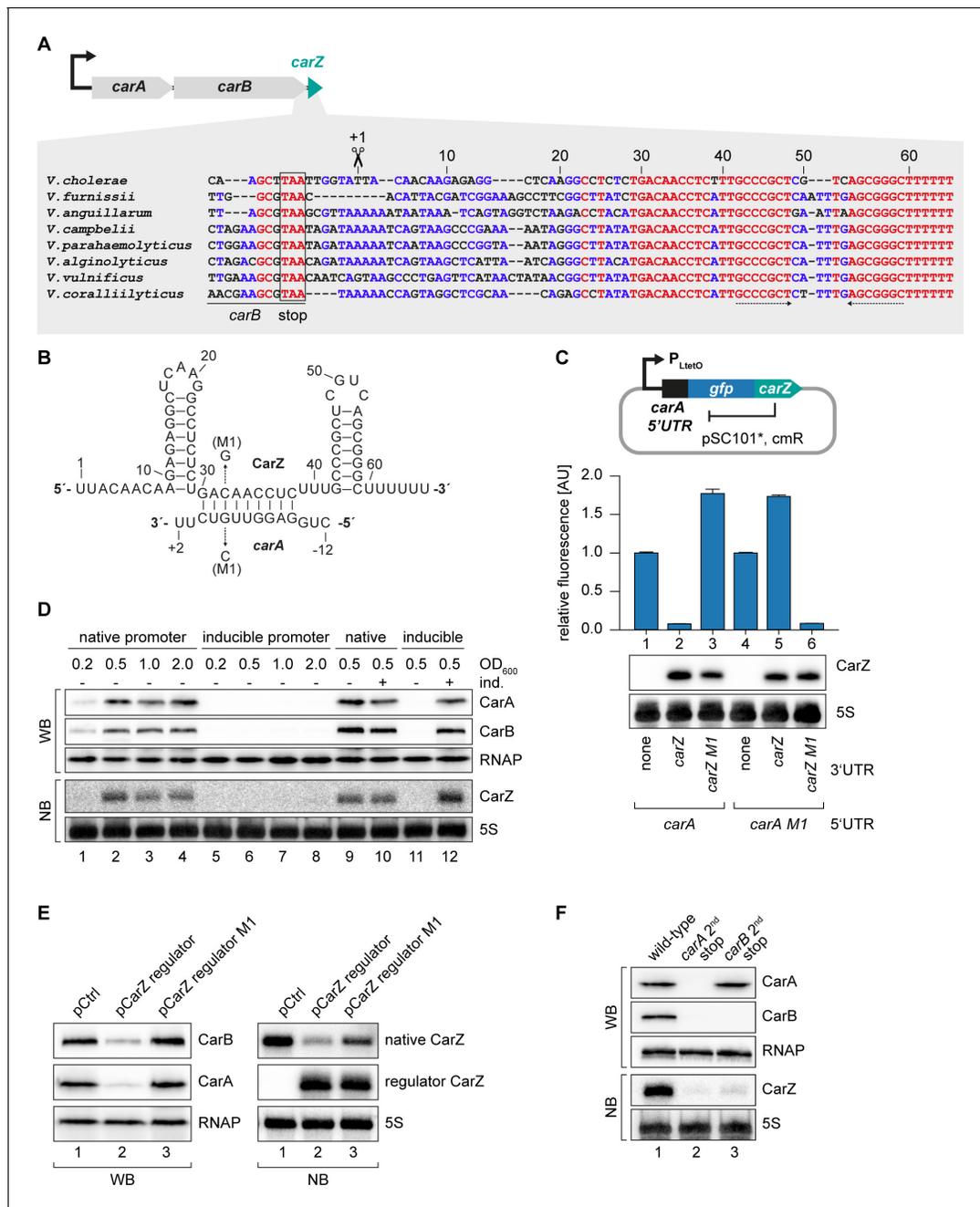
The online version of this article includes the following source data for figure 6:

**Source data 1.** Full Northern and Western blot images for the corresponding detail sections shown in **Figure 6**.

*carA* gene abrogated CarZ expression and similar results were obtained when the STOP codon was placed at the 2<sup>nd</sup> codon of *carB* (**Figure 7F**). Of note, inactivation of *carA* translation also blocked CarB production indicating, among other possibilities, that translation of the two ORFs might be coupled and that expression of CarZ relies on active translation of both ORFs. Together, these results provide evidence that CarZ is an autoregulatory sRNA and suggest that this function might be more wide-spread among the growing class of 3' UTR-derived sRNAs.

### Autoregulatory sRNAs modify the kinetics of gene induction

Bacterial sRNAs acting at the post-transcriptional level have recently been reported to add unique features to gene regulatory circuits, including the ability to promote discoordinate operon expression (*Nitzan et al., 2017*). Plasmid-borne over-expression of OppZ resulted in decreased expression of the *oppBCDF* cistrons, while leaving *oppA* levels unaffected (**Figure 3—figure supplement 1B–C**). We therefore asked if OppZ expression had a similar effect on the production of their corresponding proteins. To this end, we cultivated wild-type and *oppZ*-deficient *V. cholerae* (both carrying a control plasmid), as well as  $\Delta$ *oppZ* cells carrying an OppZ over-expression plasmid, to various stages of growth and monitored OppA and OppB levels on Western Blot (**Figure 8—figure supplement 1A**). Quantification of the results revealed a moderate increase in OppB expression (~1.8 fold) in cells lacking *oppZ* and ~5 fold decreased OppB levels when OppZ was over-expressed. Neither lack of *oppZ*, nor OppZ over-expression significantly affected OppA production (**Figure 8—figure supplement 1B–C**).



**Figure 7.** CarZ is another autoregulatory sRNA from *V. cholerae*. (A) Top: Genomic context of *carAB* and *carZ*. Bottom: Alignment of *carZ* sequences, including the last codons of *carB*, from various *Vibrio* species. The *carB* stop codon, the RNase E cleavage site and the Rho-independent terminator are indicated. (B) Predicted CarZ secondary structure and base-pairing to *carA*. Arrows indicate the single nucleotide mutations tested in (C). (C) Single-plasmid feedback regulation of *carA* by CarZ was measured by inserting the indicated *carZ* variant into the 3' UTR of a translational *carA*::*gfp* fusion. Expression was driven from a constitutive promoter. *E. coli* strains carrying the respective plasmids were grown to OD<sub>600</sub> = 1.0 and GFP production was measured. Fluorophore levels from control fusions without an sRNA gene were set to one and error bars represent the SD of three biological replicates. CarZ expression was tested by Northern blot; 5S rRNA served as loading control. (D) Protein and RNA samples were obtained from *V. cholerae* *carA*::3XFLAG *carB*::3XFLAG carrying either the native *carA* promoter or the inducible pBAD promoter upstream of *carA*. Samples were collected at the indicated OD<sub>600</sub> and tested for CarA and CarB production by Western blot and for CarZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. Lanes 1–8: Growth without L-arabinose. Lanes 9–12: Growth with either H<sub>2</sub>O (-) or L-arabinose (+) (0.2% final conc.). (E) *V. cholerae* *carA*::3XFLAG *carB*::3XFLAG strains carrying a control plasmid or a plasmid expressing a CarZ variant with a mutated stem loop (*regulator* CarZ) were grown to late exponential phase (OD<sub>600</sub> of 1.0). CarA and CarB production were tested by Western blot and expression of native CarZ or regulator CarZ was monitored on Northern blot using oligonucleotides binding to the respective loop sequence

Figure 7 continued on next page

Figure 7 continued

variants. RNAP and 5S rRNA served as loading controls for Western blot and Northern blot, respectively. (F) *V. cholerae* *carA*::3XFLAG *carB*::3XFLAG strains with the following *carA* or *carB* mutations were grown: wild-type (lane 1) or a STOP codon inserted at the 2<sup>nd</sup> codon of *carA* (lane 2) or *carB* (lane 3), respectively. At late exponential phase (OD<sub>600</sub> of 1.0), protein and RNA samples were collected and tested for CarA and CarB production by Western blot and for CarZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. The online version of this article includes the following source data and figure supplement(s) for figure 7:

**Source data 1.** Full blot images for the corresponding detail sections shown in **Figure 7** and raw data for fluorescence measurements.

**Figure supplement 1.** Hfq-dependent, post-transcriptional repression of CarA and CarB by CarZ.

**Figure supplement 1—source data 1.** Full Northern blot images for the corresponding detail sections shown in **Figure 4—figure supplement 1** and raw data for fluorescence measurements.

**Figure supplement 2.** CarZ induces *carAB* degradation.

**Figure supplement 2—source data 1.** Raw data for transcript changes as determined by qRT-PCR.

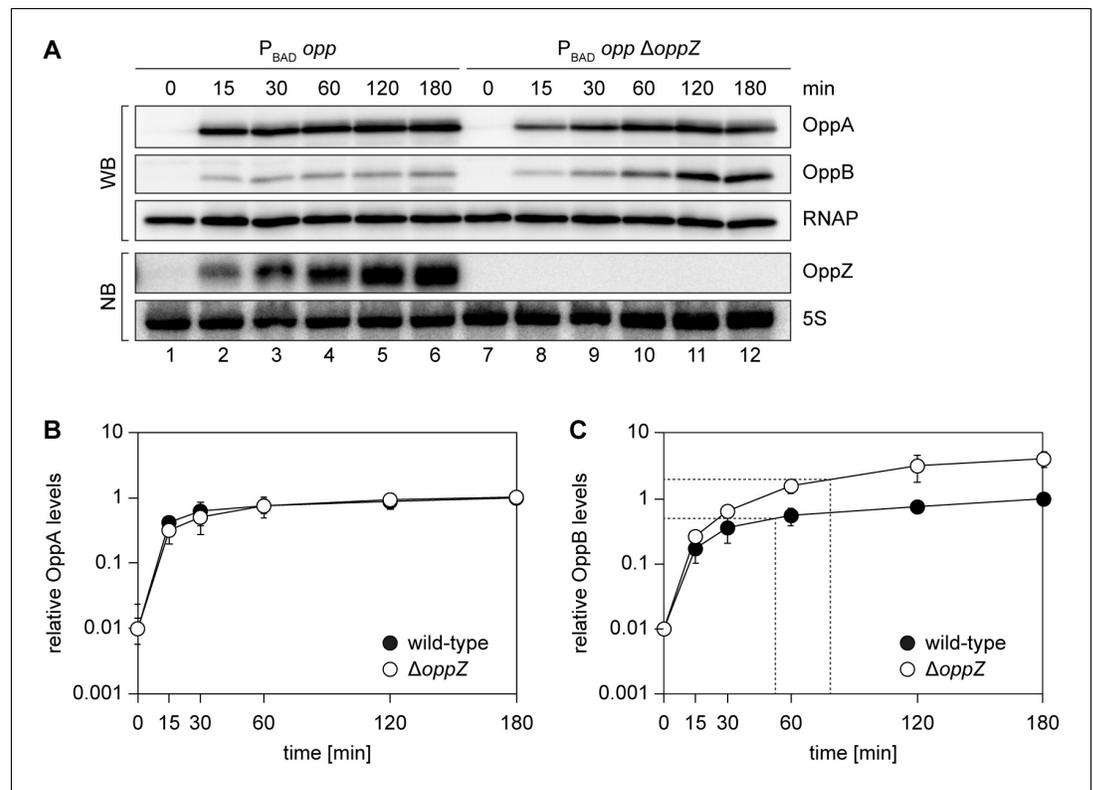
Given the relatively mild effect of *oppZ* deficiency on steady-state OppB protein levels (**Figure 8—figure supplement 1A**), we next investigated the role of OppZ on the dynamics of OppABCDF expression. Specifically, transcription factor-controlled negative autoregulation has been reported to affect the response time of regulatory networks (*Rosenfeld et al., 2002*) and we speculated that sRNA-mediated feedback control could have a similar effect. To test this hypothesis, we employed a *V. cholerae* strain in which we replaced the native promoter upstream of the chromosomal *oppA* gene with the L-arabinose-inducible pBAD promoter (see **Figure 2C**) and monitored the kinetics of OppA and OppB production in wild-type and  $\Delta oppZ$  cells before and at several time-points post induction (**Figure 8A**). Whereas OppA protein accumulated equally in wild-type and *oppZ* mutants (**Figure 8B**), expression of OppB was significantly increased in  $\Delta oppZ$  cells (**Figure 8C**). This effect was most prominent at later stages after induction (>30 min) and coincided with accumulation of OppZ (**Figure 8A**). Calculation of the OppB response time (50% of the maximal expression value) showed a significant delay in  $\Delta oppZ$  cells (~78 min), when compared to the wild-type control (~52 min). We therefore conclude that alike transcription factors, autoregulatory sRNAs change the dynamics of their associated genes, however, in contrast to transcription factors, sRNAs act at the post-transcriptional level and can direct this effect towards a specific subgroup of genes within an operon.

## Discussion

Base-pairing sRNAs regulating the expression of *trans*-encoded mRNAs are a major pillar of gene expression control in bacteria (*Gorski et al., 2017*). Transcriptomic data obtained from various microorganisms have shown that sRNAs are produced from almost all genomic loci and that the 3' UTRs of coding genes are a hotspot for sRNAs acting through Hfq (*Adams and Storz, 2020*). Expression of 3' UTR-derived sRNAs can either occur by independent promoters, or by ribonucleolytic cleavage typically involving RNase E (*Miyakoshi et al., 2015*). In the latter case, production of the sRNA is intimately connected to the activity of the promoter driving the expression of the upstream mRNA, suggesting that the regulatory function of the sRNA is linked to the biological role of the associated genes. Indeed, such functional interdependence has now been demonstrated in several cases (*Chao and Vogel, 2016; De Mets et al., 2019; Huber et al., 2020; Miyakoshi et al., 2019; Wang et al., 2020*), however, it remained unclear if and how these sRNAs also affected their own transcripts. In this regard, OppZ and CarZ provide a paradigm for 3' UTR-derived sRNAs allowing autoregulation at the post-transcriptional level. This new type of feedback inhibition is independent of auxiliary transcription factors and we could show that autoregulation by sRNAs can either involve the full transcript (CarZ), or act at the suboperonic level (OppZ).

## Features of RNase E-mediated gene control

RNase E is a principal factor for RNA turnover in almost all Gram-negative bacteria (*Bandyra and Luisi, 2018*). The protein forms a tetramer in vivo and serves as the scaffold for the degradosome, a large, multi-enzyme complex typically containing the phosphorolytic exoribonuclease PNPase, the RNA-helicase RhlB, and the glycolytic enzyme enolase (*Ait-Bara and Carpousis, 2015*). Substrates of RNase E are preferentially AU-rich and harbor a 5' mono-phosphate. Thus, the enzyme relies on



**Figure 8.** Modified kinetics of gene induction by autoregulatory OppZ. (A) Expression of the *opp* operon including the *oppA*::3XFLAG and *oppB*::3XFLAG genes and the native *oppZ* gene (lanes 1–6) or an *oppZ* deletion (lanes 7–12) was induced from the pBAD promoter at late exponential phase ( $OD_{600}$  of 1.0) by the addition of L-arabinose (0.2% final conc.). Protein and RNA samples were obtained at the indicated time points and tested for OppA and OppB production by Western blot and for OppZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. (B, C) Quantification of OppA (B) or OppB (C) levels from the experiment in (A); error bars represent the SD of three biological replicates. Data are presented as fold regulation of OppA or OppB in  $\Delta oppZ$  compared to the wild-type. Dashed lines in (C) indicate the time points of half-maximum OppB expression.

The online version of this article includes the following source data and figure supplement(s) for figure 8:

**Source data 1.** Quantification of OppAB protein levels from Western blots and full blot images for the corresponding detail sections shown in **Figure 8**.

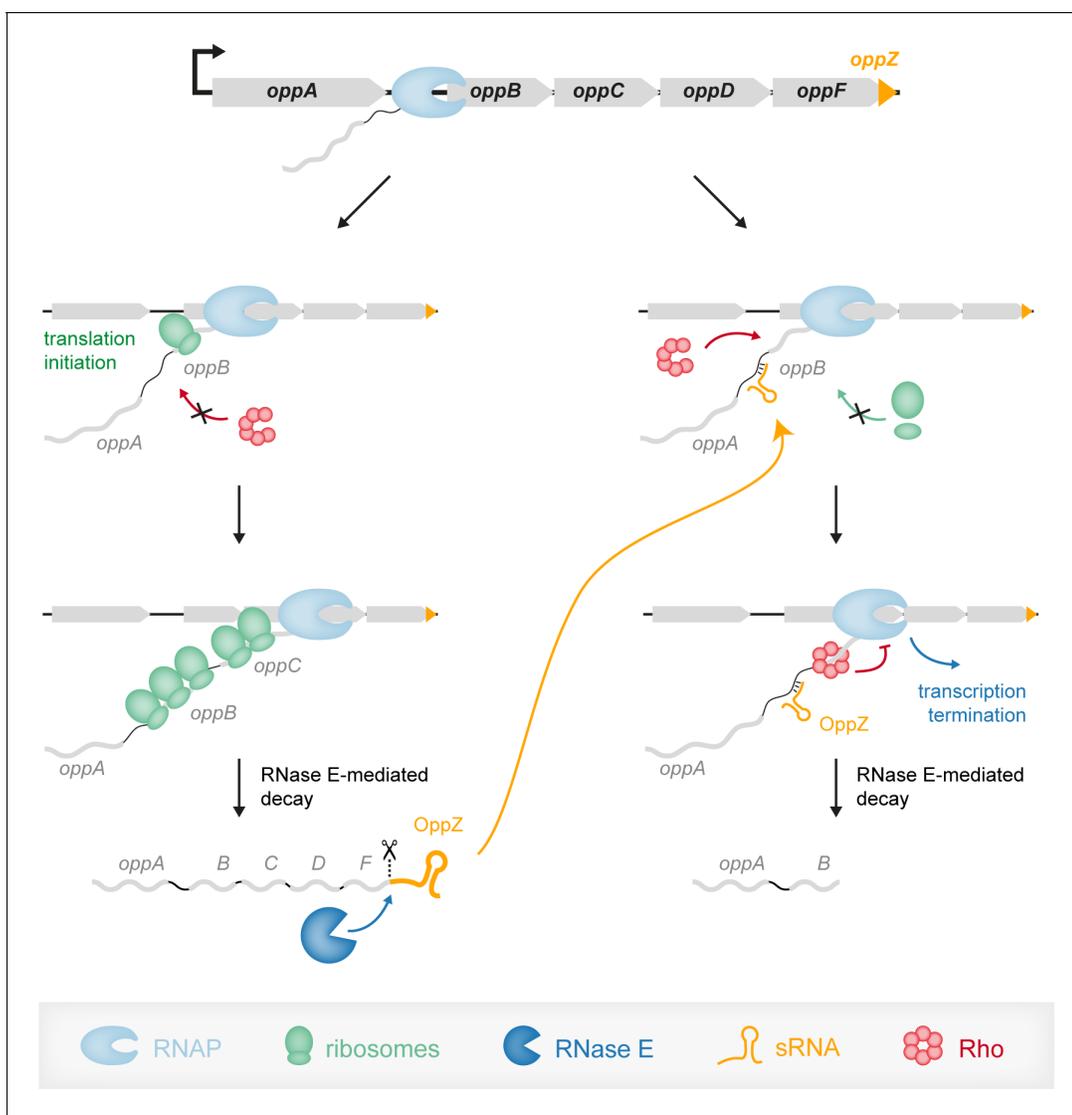
**Figure supplement 1.** OppZ-dependent repression of OppA and OppB protein levels.

**Figure supplement 1—source data 1.** Quantification of OppAB protein levels from Western blots and full blot images for the corresponding detail sections shown in **Figure 8—figure supplement 1**.

RNA pyrophosphohydrolases such as RppH, which convert the 5' terminus from a triphosphate to a monophosphate, before transcript degradation can be initiated (Deana *et al.*, 2008). Recognition of a substrate is followed by scanning of RNase E for suitable cleavage sites along the transcript (Richards and Belasco, 2019). TIER-seq-based identification of a consensus sequence for RNase E target recognition revealed highly similar motifs for *V. cholerae* (Figure 1D) and *S. enterica* (Chao *et al.*, 2017). These results further support the previously proposed 'U<sub>+2</sub> Ruler-and-Cut' mechanism, in which a conserved uridine located two nts down-stream of the cleavage site is key for RNase E activity. However, in contrast to the data obtained from *S. enterica*, we discovered only a mild enrichment of RNase E cleavage sites occurring at translational stop codons (Figure 1—figure supplement 3A). This observation might be explained by differences in stop codon usage between *V. cholerae* and *S. enterica* (Korkmaz *et al.*, 2014) and could point to species-specific features of RNase E activity.

## The role of termination factor Rho in sRNA-mediated gene expression control

Approximately 25–30% of all genes in *E. coli* depend on Rho for transcription termination (Cardinale et al., 2008; Dar and Sorek, 2018b; Peters et al., 2012). BCM treatment of *V. cholerae* wild-type cells revealed 699 differentially regulated genes (549 upregulated and 150 repressed genes; Supplementary file 3A), suggesting an equally global role for Rho in this organism. Rho-dependent transcription termination is modulated by various additional factors (Mitra et al., 2017). This includes anti-termination factors such as NusG, as well as Hfq and its associated sRNAs (Bossi et al., 2020). For sRNAs, the effect on Rho activity can be either activating or repressing. Previous work has shown that sRNAs can mask Rho-dependent termination sites and thereby promote transcriptional read-through (Lin et al., 2019; Sedlyarova et al., 2016). Negative gene regulation involving sRNAs and Rho typically includes translation inhibition by the sRNA resulting in separation of transcription and translation complexes (Figure 9). Coupling of transcription and translation normally protects the nascent mRNA from Rho action and loss of ribosome binding supports



**Figure 9.** Model of the OppZ-dependent mechanism of *opp* regulation. Transcription of the *oppABCDF* operon initiates upstream of *oppA* and in the absence of OppZ (left) involves all genes of the operon as well as OppZ. In this scenario, all cistrons of the operon are translated. In the presence of OppZ (right), the sRNA blocks translation of *oppB* and the ribosome-free mRNA is recognized by termination factor Rho. Rho catches up with the transcribing RNAP and terminates transcription pre-maturely within *oppB*. Consequently, *oppBCDF* are not translated and OppZ is not produced.

transcription termination (Bossi et al., 2012). In addition, lack of ribosome-mediated protection can render the mRNA target vulnerable to ribonucleases, e.g. RNase E, which can also lead to the degradation of the sRNA (Feng et al., 2015; Massé et al., 2003). Which of these mechanisms are at play for a given sRNA-target mRNA pair is most often unknown and it is likely that both types of regulation can occur either independently or in concert. For example, over-expression of OppZ did not affect *oppB* transcript stability (Figure 3—figure supplement 1D), suggesting that induction of Rho-mediated transcription termination is the main mechanism for gene repression in this sRNA-target mRNA pair. In contrast, analogous experiments testing the stability of the *carA* and *carB* transcripts upon CarZ over-expression revealed a significant drop in transcript stability for both mRNAs (Figure 7—figure supplement 2A–B). These results suggest that translation inhibition of *carA* by CarZ has two outcomes: 1<sup>st</sup>) accelerated ribonucleolytic decay of the *carAB* transcript and 2<sup>nd</sup>) Rho-mediated transcription termination. Using two regulatory mechanisms (CarZ-*carA*) instead of one (OppZ-*oppB*) might explain the strong inhibition of *carA::gfp* by CarZ (~10 fold, Figure 7C), when compared to the relatively weak repression (1.8-fold) of *oppB::gfp* by OppZ (Figure 3D).

Employing multiple regulatory mechanisms on one target mRNA might have led to an underestimation of the prevalence of Rho-mediated transcription termination in sRNA-mediated gene control. In fact, sRNAs frequently repress genes that are downstream in an operon with their base-pairing target, which could point to a possible involvement of Rho (Bossi et al., 2020). Rho is known to bind cytosine-rich RNA elements (Allfano et al., 1991), however, due to the strong variability in size and composition of these sequences, predicting Rho binding sites (a.k.a. *rut* sites) from genomic or transcriptomic data has been a difficult task (Nadiras et al., 2018). Indeed, while our transcriptomic data of the *oppB* start codon mutant did not allow us to pinpoint the position of the *rut* site in *oppB* (Figure 5C), evidence obtained from genetic analyses using various *oppB* STOP codon mutants revealed that Rho-dependent termination likely occurs at or close to codon 115 in *oppB* (Figure 6B). We attribute the lack of this termination event in the transcriptomic data to the activity of 3'–5' acting exoribonucleases (e.g. RNase II or PNPase Bechhofer and Deutscher, 2019; Mohanty and Kushner, 2018), which degrade the untranslated *oppB* sequence. Identifying the relevant exonucleases might well allow for an advanced annotation of global Rho-dependent termination sites and cross-comparison with documented sRNA-target interaction could help to clarify the relevance of Rho-mediated termination in sRNA-based gene control.

## Dynamics of RNA-based feedback regulation

Transcription factors and sRNAs are the principal components of gene networks. While the regulatory outcome of sRNA and transcription factor activity is often very similar, the underlying regulatory dynamics are not (Hussein and Lim, 2012). Regulatory networks involving sRNAs and transcription factors are called mixed circuits and have now been studied in greater detail. Similar to systems relying on transcription factors, feedback regulation is common among sRNAs (Nitzan et al., 2017). However, unlike the examples presented in this study, these circuits always involve the action of a transcription factor, which has implications for their regulatory dynamics. For example, the OmpR transcription factor activates the expression of the OmrA/B sRNAs, which repress their own synthesis by inhibiting the *ompR-envZ* mRNA (Guillier and Gottesman, 2008). This constitutes an autoregulatory loop, however, given that transcription of OmrA/B ultimately relies on OmpR protein levels, this regulation will only become effective when sufficient OmpR turn-over has been achieved (Brosse et al., 2016). In contrast, autoregulatory circuits involving 3' UTR-derived sRNAs are independent of such auxiliary factors and therefore provide a more rapid response. In case of OppZ-*oppB*, we showed that the sRNA has a rapid effect on OppB expression levels (Figure 8C) and given the involvement of Rho-mediated transcription termination in this process, we expect similar dynamics for OppZ autoregulation (Figure 9).

Another key difference between feedback regulation by transcription factors and 3' UTR-derived sRNAs is the stoichiometry of the players involved. In transcription factor-based feedback loops, the mRNA coding for the autoregulatory transcription factor can go through multiple rounds of translation, which will lead to an excess of the regulator over the target promoter. The degree of autoregulation is then determined by the cellular concentration of the transcription factor and the affinity towards its own promoter (Rosenfeld et al., 2002). In contrast, autoregulatory sRNAs which are generated by ribonucleolytic cleavage come at a 1:1 stoichiometry with their targets. However, this situation changes when the sRNA controls multiple targets. For OppZ, we have shown that

*oppBCDF* is the only transcript regulated by the sRNA (**Figure 3A**) and we currently do not know if CarZ has additional targets besides *carAB*. In addition, not all sRNA-target interactions result in changes in transcript levels as previously reported for the interaction of the Qrr sRNAs with the *luxO* transcript (**Feng et al., 2015**). New technologies, for example RIL-Seq (**Melamed et al., 2020; Melamed et al., 2016**), capturing the global interactome of base-pairing sRNAs independent of their regulatory state could help to address this question and clarify the stoichiometric requirements for sRNA-mediated autoregulation.

### Possible biological relevance of autoregulatory sRNAs

Autoregulation by 3' UTR-derived sRNAs allows for discoordinate operon expression, which is in contrast to their transcription factor counterparts. This feature might be particularly relevant for long mRNAs containing multiple cistrons, such as *oppABCDF*. The *oppABCDF* genes encode an ABC transporter allowing high affinity oligopeptide uptake (**Hiles et al., 1987**). OppBCDF constitute the membrane-bound, structural components of the transport system, whereas OppA functions as a periplasmic binding protein. The overall structure of the transporter requires each one unit of OppB, OppC, OppD, and OppF, while OppA does constitutively interact with the complex and typically accumulates to higher concentrations in the periplasm (**Doeven et al., 2004**). Given that transcription of *oppABCDF* is controlled exclusively upstream of *oppA* (**Figure 2C** and **Papenfort et al., 2015b**), OppZ-mediated autoregulation of *oppBCDF* (rather than the full operon) might help to achieve equimolar concentrations of OppB, OppC, OppD, and OppF in the cell without affecting OppA production.

The *carAB* genes, which are repressed by CarZ, encode carbamoyl phosphate synthetase; an enzyme complex catalyzing the first step in the separate biosynthetic pathways for the production of arginine, and pyrimidine nucleotides (**Castellana et al., 2014**). Similar to OppBCDF, the CarAB complex contains one subunit of CarA and one subunit of CarB. Transcriptional control of *carAB* is complex and controlled by several transcription factors integrating information from purine, pyrimidine, and arginine pathways (**Charlier et al., 2018**). While the exact biological role of CarZ-mediated feedback regulation of *carAB* requires further investigation, transcription factor-based feedback regulation has been reported to reduce transcriptional noise (**Alon, 2007**), which could also be an important feature of sRNA-mediated autoregulation. The OppZ and CarZ sRNAs identified in this study now provide the framework to test this prediction.

### Orthogonal use of gene autoregulation by 3' UTR-derived sRNAs

Regulatory RNAs have now been established as powerful components of the synthetic biology toolbox (**Qi and Arkin, 2014**). RNA regulators are modular, versatile, highly programmable, and therefore ideal candidates for synthetic biology approaches. Similarly, autoregulatory loops using transcriptional repressors find ample use in synthetic regulatory circuits (**Afroz and Beisel, 2013**). While it might be counterintuitive for a transcript to also produce its own repressor, negative feedback regulation has been reported to endow regulatory networks with improved robustness when disturbances to the system are imposed. Hfq-binding sRNAs providing feedback control have recently also been demonstrated to efficiently replace transcriptional regulation in artificial genetic circuits (**Kelly et al., 2018**). However, these sRNAs were produced from separate genes and therefore required additional transcriptional input, which increases noise. In contrast, the autoregulatory sRNAs presented here are produced by ribonucleolytic cleavage and we have shown that both OppZ and CarZ are efficiently clipped off from foreign genes, such as *gfp* (**Figure 3—figure supplement 3, Figure 7C**). We therefore propose that autoregulatory sRNAs can be attached to the 3' UTR of other genes as well, offering a simple and highly modular concept to introduce autoregulation into a biological system. These circuits can be further tuned by modifying the base-pairing strength of the RNA duplex formed between the sRNA and the target, as well as the introduction of Rho-dependent termination events. The latter could be used to avoid over-production of the sRNA, which will further shape the regulatory dynamics of the system. Given that transcriptomic analyses have revealed thousands of stable 3' UTR RNA tails derived from human transcripts (**Gruber and Zavalan, 2019; Malka et al., 2017**), we believe that RNA-based gene autoregulation also could be present and find applications in higher organisms.

## Materials and methods

## Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background ( <i>Escherichia coli</i> )	See <a href="#">Supplementary file 4</a>	This study		See <a href="#">Supplementary file 4</a>
Strain, strain background ( <i>Vibrio cholerae</i> )	See <a href="#">Supplementary file 4</a>	This study		See <a href="#">Supplementary file 4</a>
Recombinant DNA reagent (plasmids)	See <a href="#">Supplementary file 4</a>	This study		See <a href="#">Supplementary file 4</a>
Sequence-based reagent (oligonucleotides)	See <a href="#">Supplementary file 4</a>	This study		See <a href="#">Supplementary file 4</a>
Antibody	ANTI-FLAG M2 antibody (mouse monoclonal)	Sigma-Aldrich	Cat#F1804; RRID:AB_262044	(Western blot 1:1.000)
Antibody	RNA Polymerase alpha antibody 4RA2 (rabbit monoclonal)	BioLegend	Cat#WP003; RRID:AB_2687386	(1:10.000)
Antibody	anti-mouse IgG HRP (goat polyclonal)	ThermoFischer	Cat#31430; RRID:AB_228307	(1:10.000)
Antibody	anti-rabbit IgG HRP (goat polyclonal)	ThermoFischer	Cat#A16104; RRID:AB_2534776	(1:10.000)
Commercial assay or kit	TURBO DNA-free Kit	Invitrogen	Cat#AM1907	
Commercial assay or kit	NEBNext Ultra II Directional RNA Library Prep Kit for Illumina	NEB	Cat#E7760	
Commercial assay or kit	Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria)	Illumina	Cat#MRZGN126	
Chemical compound, drug	Protein G Sepharose	Sigma-Aldrich	Cat##P3296	
Chemical compound, drug	Bicyclomycin (BCM)	SantaCruz Biotech.	Cat#sc-391755; CAS ID: 38129-37-2	
Software, algorithm	MultAlin	<b>Corpet, 1988</b> (PMID:2849754)		<a href="http://multalin.toulouse.inra.fr/multalin">http://multalin.toulouse.inra.fr/multalin</a>
Software, algorithm	RNAhybrid	<b>Rehmsmeier et al., 2004</b> (PMID:15383676)		<a href="http://bibiserv2.cebitec.uni-bielefeld.de">http://bibiserv2.cebitec.uni-bielefeld.de</a> RRID:SCR_003252
Software, algorithm	CLC Genomics Workbench	Qiagen		<a href="https://qiagenbioinformatics.com">https://qiagenbioinformatics.com</a> RRID:SCR_011853
Software, algorithm	SigmaPlot	SYSTAT		<a href="https://systatsoftware.com">https://systatsoftware.com</a> RRID:SCR_003210

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Software, algorithm	GelQuantNET	biochemlabsolutions		<a href="http://biochemlabsolutions.com/GelQuantNET.html">http://biochemlabsolutions.com/GelQuantNET.html</a> RRID:SCR_015703
Software, algorithm	BIO-1D	VILBER		<a href="http://vilber.de/en/products/analysis-software">http://vilber.de/en/products/analysis-software</a>
Software, algorithm	ImageJ	<b>Schneider et al., 2012</b> (PMID:22930834)		<a href="https://imagej.nih.gov/ij/">https://imagej.nih.gov/ij/</a> RRID:SCR_003070
Software, algorithm	cutadapt	<b>Martin, 2011</b>		<a href="https://doi.org/10.14806/ej.17.1.200">https://doi.org/10.14806/ej.17.1.200</a>
Software, algorithm	READemption	<b>Förstner et al., 2014</b> (PMID:25123900)		<a href="https://doi.org/10.5281/zenodo.591469">https://doi.org/10.5281/zenodo.591469</a>
Software, algorithm	DESeq2	<b>Love et al., 2014</b> (PMID:25516281)		<a href="http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html">http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
Software, algorithm	RNAfold	<b>Lorenz et al., 2011</b> (PMID:22115189)		<a href="http://www.tbi.univie.ac.at/RNA">http://www.tbi.univie.ac.at/RNA</a>
Software, algorithm	WebLogo	<b>Crooks et al., 2004</b> (PMID:15173120)		<a href="http://weblogo.threeplusone.com/">http://weblogo.threeplusone.com/</a>
Software, algorithm	BEDTools	<b>Quinlan and Hall, 2010</b> (PMID:20110278)		<a href="http://code.google.com/p/bedtools">http://code.google.com/p/bedtools</a>

### Strains, plasmids, and growth conditions

Bacterial strains, plasmids and DNA oligonucleotides used in this study are listed in **Supplementary file 4**. Throughout the study, *V. cholerae* C6706 (**Thelin and Taylor, 1996**) was used as the wild-type strain. *V. cholerae* and *E. coli* strains were grown aerobically in LB medium at 37°C except for temperature-sensitive strains. For stationary phase cultures of *V. cholerae*, samples were collected with respect to the time point when the cells reached an OD<sub>600</sub> >2.0, i.e., 3 hr after cells reached an OD<sub>600</sub> reading of 2.0. For transcript stability experiments, rifampicin was used at 250 µg/ml. To inhibit Rho-dependent transcription termination, bicyclomycin (BCM; sc-391755; Santa Cruz Biotechnology, Dallas, Texas) was used at 25 µg/ml. Other antibiotics were used at the following concentrations: 100 µg/ml ampicillin; 20 µg/ml chloramphenicol; 50 µg/ml kanamycin; 50 U/ml polymyxin B; and 5,000 µg/ml streptomycin.

For transient inactivation of RNase E, *V. cholerae* wild-type and a temperature-sensitive strain harboring the *rne-3071* mutation were grown at 30°C to the indicated cell density. Cultures were divided in half and either continuously grown at 30°C or shifted to 44°C. RNA samples were collected from both strains and temperatures at the indicated time points after the temperature shift.

RK2/RP4-based conjugal transfer was used to introduce plasmids into *V. cholerae* from *E. coli* S17λpir plasmid donor strains (**Simon et al., 1983**). Subsequently, transconjugants were selected using appropriate antibiotics and polymyxin B to specifically inhibit *E. coli* growth. *V. cholerae* mutant strains were generated as described previously (**Papenfort et al., 2015b**). Briefly, pKAS32 plasmids were transferred into *V. cholerae* strains by conjugation and cells were screened for ampicillin resistance. Single colonies were streaked on streptomycin plates for counter-selection and colonies were tested for desired mutations by PCR or sequencing. Strain KPEC53467 was generated by phage P1 transduction to transfer the Δ*hfq*::KanR allele (**Baba et al., 2006**) into *E. coli* Top 10 and subsequent removal of the KanR cassette using plasmid pCP20 **Datsenko and Wanner, 2000** following standard protocols.

## Plasmid construction

The plasmids used in this study are listed in **Supplementary file 4B**, used DNA oligonucleotides are listed in **Supplementary file 4C**. For pMD004, the *rrnB* terminator from pKP8-35 (**Papenfort et al., 2015b**) was amplified with KPO-1484/1485 and cloned by Gibson assembly into pKP-331 (**Papenfort et al., 2015b**) linearized with KPO-0196/1397. pMD089 was generated by amplification of *oppZ* from KPS-0014 chromosomal DNA using KPO-2552/2553 and Gibson assembly with pMD004 linearized with KPO-0196/1397. pMD373 was constructed by amplification of *oppB::3XFlag* from KPVC11709 chromosomal DNA using KPO-5878/5879 and Gibson assembly with pMD004 linearized with KPO-2789/pBAD-ATGrev. pCMW-2 was obtained by removing the promoterless *gfp* from pCMW-1 (**Waters and Bassler, 2006**) by amplification with KPO-2757/5421. pMD090 was generated by amplification of *oppZ* from KPS-0014 chromosomal DNA using KPO-2568/2553 and Gibson assembly with pEVS143 (**Dunn et al., 2006**) linearized with KPO-0092/1397. The M1 point mutation was introduced into pMD090 by site-directed mutagenesis with KPO-2619/2620, yielding pMD118. pMD194 and pMD195 were obtained by site-directed mutagenesis of pMD090 and pMD118, respectively, with KPO-3190/3191. pMD397 and pMD398 were obtained by replacing the p15a origin of replication in pCMW-1 and pMD194, respectively, by the pSC101 origin including an E93K mutation in the *repA* sequence. To this end, pCMW-1 and pMD194 were linearized with KPO-2041/2049, the pSC101 origin was amplified from pXG10-SF (**Corcoran et al., 2012**) in three parts (with KPO-6490/6493, KPO-6492/6495 and KPO-6494/6491) and fragments were joined with Gibson assembly. pMD173 and pMD174 were generated by amplification of the pBR322 origin from pBAD-Myc-His (Invitrogen) with KPO-2042/2043 and Gibson assembly with pCMW-1 or pMD090, respectively (both linearized with KPO-2041/2049). pMD197 was obtained by replacing the *oppZ* gene in pMD174 with a longer *oppF-oppZ* fragment (amplified from KPS-0014 chromosomal DNA using KPO-3197/2553) by Gibson assembly. pNP015 was constructed by amplification of *carZ* from KPS-0014 chromosomal DNA using KPO-1013/1014 and subcloning into linearized pEVS143 (KPO-0092/1023) with XbaI. Again, the M1 point mutation was introduced into pNP015 by site-directed mutagenesis with KPO-1782/1783, yielding pMH013. pMD361 and pMD362 were obtained by site-directed mutagenesis of pNP015 and pMH013, respectively, with KPO-5686/5687.

For translational GFP reporters, pMD093 was generated by amplification of the *oppAB* intergenic region and the first 5 codons of *oppB* from KPS-0014 chromosomal DNA using KPO-2580/2583 and Gibson assembly with pXG10-SF linearized with KPO-1702/1703. Site-directed mutagenesis of pMD093 with KPO-2615/2616 yielded pMD125. Accordingly, pMH010 and pMD374 were generated by amplification of the *carA* 5'UTR and the first 20 codons of *carA* with KPO-1674/1675 (for pMH010) or a fragment including the *carA* 5' UTR, the complete *carA* gene and the first 20 codons of *carB* with KPO-1674/5874 (for pMD374) from KPS-0014 chromosomal DNA, followed by Gibson assembly with pXG10-SF linearized with KPO-1702/1703. Site-directed mutagenesis of pMH010 and pMD374 with KPO-1778/1779 yielded pMH012 and pMD375, respectively. For discoordinate translational reporters for *oppB* to *oppF*, fragments from the *oppAB* intergenic region to the first 5 codons of *oppB* or the first 20 codons of *oppC*, *oppD* or *oppF* were amplified from KPS-0014 chromosomal DNA using KPO-2622 and KPO-2583 (*oppB*), KPO-2577 (*oppC*), KPO-2578 (*oppD*) or KPO-2579 (*oppF*). *mKate2* was amplified from pMD079 (**Herzog et al., 2019**) with KPO-2511/2625 and the pXG10-SF backbone was linearized with KPO-2621/1703. Gibson assembly was used to join the pXG10-SF backbone, *mKate2* and the respective *opp* fragment to generate pMD120, pMD352, pMD353 and pMD354. Site-directed mutagenesis of pMD120 and pMD354 with KPO-2615/2616 yielded pMD129 and pMD355, respectively.

pMD091 and pMD112 were constructed by amplification of *oppZ* from KPS-0014 chromosomal DNA using KPO-2585/2586 and Gibson assembly with pXG10-SF (for pMD091) or pMD093 (for pMD112), both linearized with KPO-2584/2508. The M1 mutations in the *oppAB* IGR or *oppZ* were obtained by site-directed mutagenesis of pMD112 with KPO-2615/2616 or KPO-2617/2618, respectively, to construct pMD117, pMD127 and pMD128. Site-directed mutagenesis of pMD091 and pMD093 with KPO-2665/2666 to introduce the M2 mutation into *oppZ* yielded pMD124 and pMD126, respectively. Accordingly, pMD294 and pMD297 were constructed by amplification of *carZ* from KPS-0014 chromosomal DNA using KPO-4815/4817 and Gibson assembly with pMH010 (for pMD294) or pMH012 (for pMD297), both linearized with KPO-2584/2508. Site-directed mutagenesis of pMD294 and pMD297 with KPO-1782/1783 yielded pMD296 and pMD298, respectively.

All pKAS32-derived plasmids (*Skorupski and Taylor, 1996*) were constructed by Gibson assembly of the respective up and down flanks with the pKAS32 backbone (linearized with KPO-0267/0268) and an additional fragment containing the 3XFLAG sequence or an *araC*-pBAD fragment where appropriate. Flanks were amplified from KPS-0014 chromosomal DNA unless otherwise stated. Plasmids for gene deletions or chromosomal point mutations are listed in the following with the respective primer pairs for up and down flanks indicated: pMD003 (KPO-1440/1443 and KPO-1441/1442), pMD160 (KPO-2753/1199 and KPO-1200/2754), pMD350 (KPO-1429/1289 and KPO-1290/1430), pMD349 (KPO-5243/5244 from KPVC11709 chromosomal DNA and KPO-5245/5246), pMD357 (KPO-5243/5672 and KPO-5673/5246, both from KPVC11709 chromosomal DNA), pMD358 (KPO-5243/5674 and KPO-5675/5246, both from KPVC11709 chromosomal DNA), pMD370 (KPO-5880/5884 and KPO-5885/5881, both from KPVC11709 chromosomal DNA), pMD371 (KPO-5880/5886 and KPO-5887/5881, both from KPVC11709 chromosomal DNA), pMD372 (KPO-5882/5890 and KPO-5891/5883, both from KPVC11709 chromosomal DNA), pMD356 (KPO-3183/5670 and KPO-5671/3186, both from KPVC11709 chromosomal DNA), pMD367 (KPO-4395/5824 from KPVC11709 chromosomal DNA and KPO-5823/4400), pMD369 (KPO-4379/5828 and KPO-5827/4384), pMD385 (KPO-5235/6029 and KPO-6030/5238, both from KPVC12872 chromosomal DNA) and pMD386 (KPO-5223/6031 and KPO-6032/5226, both from KPVC12872 chromosomal DNA). For pMD199 and pMD200, flanks were amplified with KPO-3179/3180 and KPO-3181/3182 (for pMD199) or with KPO-3183/3184 and KPO-3185/3186 (for pMD200). The 3XFLAG fragment was obtained by annealing of the oligonucleotides KPO-3157/3158. Flanks and 3XFLAG tag for pMD269, pMD346 and pMD347 were amplified with the following oligonucleotides: KPO-4385/4386, KPO-4387/4388 and KPO-4389/4390 (for pMD269); KPO-5223/5224, KPO-5225/5226 and KPO-5231/5232 (for pMD346); KPO-5227/5228, KPO-5229/5230 and KPO-5233/5234 (for pMD347). pMD199 was used as template for the 3XFLAG fragments. For pMD280 and pMD351, a fragment containing the *araC* gene and the pBAD promoter was amplified from pMD004 using 4529/0196. Flanks were amplified with KPO-4527/4528 and KPO-4530/4531 (for pMD280) or with KPO-5235/5236 and KPO-5237/5238 (for pMD351).

### RNA isolation, Northern blot analysis and quantitative real-time PCR

For Northern blot analyses, total RNA was prepared and blotted as described previously (*Papenfort et al., 2017*). Membranes were hybridized in Roti-Hybri-Quick buffer (Carl Roth, Karlsruhe, Germany) with [<sup>32</sup>P]-labeled DNA oligonucleotides at 42°C or with riboprobes at 63°C. Riboprobes were generated using the MAXIscript T7 Transcription Kit (Thermo Fisher Scientific, Waltham, Massachusetts). Signals were visualized using a Typhoon Phosphorimager (GE Healthcare, Chicago, Illinois) and quantified using GelQuant (RRID:SCR\_015703; BioChemLabSolutions, San Francisco, California). Oligonucleotides for Northern blot analyses are provided in **Supplementary file 4C**. For qRT-PCR, total RNA was isolated with the SV Total RNA Isolation System (Promega, Fitchburg, Wisconsin). qRT-PCR was performed in three biological and two technical replicates using the Luna Universal One-Step RT-qPCR Kit (New England BioLabs, Ipswich, Massachusetts) and the MyiQ Single-Color Real-Time PCR Detection System (Bio-Rad, Hercules, California). 5S rRNA and *recA* were used as reference genes; oligonucleotides used for all qRT-PCR analyses are provided in **Supplementary file 4C**.

### Hfq co-immunoprecipitation

Hfq co-immunoprecipitations were performed as previously described (*Huber et al., 2020*). Briefly, *V. cholerae* wild-type (KPS-0014) and *hfq::3XFLAG* (KPS-0995) (*Peschek et al., 2019*) strains were grown in LB medium to OD<sub>600</sub> of 2.0. Lysates corresponding to 50 OD<sub>600</sub> units were subjected to immunoprecipitation using monoclonal anti-FLAG antibody (#F1804; Sigma-Aldrich, St. Louis, Missouri) and Protein G Sepharose (#P3296; Sigma-Aldrich).

### Western blot analysis and fluorescence assays

Total protein sample preparation and Western blot analyses were performed as described previously (*Papenfort et al., 2017*). Signals were visualized using a Fusion FX EDGE imager (Vilber Lourmat, Marne-la-Vallée, France) and band intensities were quantified using the BIO-1D software (Vilber Lourmat). 3XFLAG-tagged fusions were detected using mouse anti-FLAG antibody (#F1804; RRID:

AB\_262044; Sigma-Aldrich) and goat anti-mouse HRP-conjugated IgG antibody, (#31430; RRID:AB\_228307; Thermo Fisher Scientific). RNAP $\alpha$  served as a loading control and was detected using rabbit anti-RNAP $\alpha$  antibody (#WP003; RRID:AB\_2687386; BioLegend, San Diego, California) and goat anti-rabbit HRP-conjugated IgG antibody, (#16104; AB\_2534776; Thermo Fisher Scientific). Fluorescence assays of *E. coli* strains to measure mKate and GFP expression were performed as previously described (Urban and Vogel, 2007). Cells were washed in PBS and fluorescence intensity was quantified using a Spark 10 M plate reader (Tecan, Männedorf, Switzerland). Control strains not expressing fluorescent proteins were used to subtract background fluorescence.

### RNA-seq analysis: TIER-seq

*V. cholerae* wild-type and *rne*<sup>TS</sup> strains were grown in biological triplicates at 30°C to OD<sub>600</sub> of 1.0. Cultures were divided in half and either continuously grown at 30°C or shifted to 44°C. Cells were harvested from both strains and temperatures at 60 min after the temperature shift by addition of 0.2 volumes of stop mix (95% ethanol, 5% (v/v) phenol) and snap-frozen in liquid nitrogen. Total RNA was isolated and digested with TURBO DNase (Thermo Fisher Scientific). cDNA libraries were prepared by vertis Biotechnology AG (Freising, Germany): total RNA samples were poly(A)-tailed and 5'PPP structures were removed using RNA 5'Polyposphatase (Epicentre, Madison, Wisconsin). An RNA adapter was ligated to the 5' monophosphate and first-strand cDNA synthesis was performed using an oligo(dT)-adapter and M-MLV reverse transcriptase. The resulting cDNAs were PCR-amplified, purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics, Chaska, Minnesota) and sequenced using a NextSeq 500 system in single-read mode for 75 cycles.

After quality trimming and adapter clipping with cutadapt (version 2.5, DOI: <https://doi.org/10.14806/ej.17.1.200>) the sequencing reads were mapped to the *V. cholerae* reference genome (NCBI accession numbers: NC\_002505.1 and NC\_002506.1) including annotations for Vcr001-Vcr107 (Papenfert et al., 2015b) using READemption's (Förstner et al., 2014, v0.5.0, <https://doi.org/10.5281/zenodo.591469>) sub-command 'align' (building on segemehl version 0.3.4, Hoffmann et al., 2009) and nucleotide-specific coverage values were calculated with the sub-command 'coverage' based on the first base of the reads. Positions with a coverage of 20 reads or more were used to perform an enrichment analysis using DESeq2 (v.1.20.0, Love et al., 2014) comparing the WT to the mutant libraries. Nucleotides for which DESeq2 calculated an absolute fold-change of 3.0 or more and an adjusted (Benjamini-Hochberg corrected) p-value of 0.05 were treated in following analysis steps as bona fide cleavage sites.

The Minimum free energy (MFE) of sequence windows was computed with RNAfold (version 2.4.14) of the Vienna package (Lorenz et al., 2011). Sequence logos were created with WebLogo (version 3.7.4; Crooks et al., 2004). Overlaps of cleavage sites with other features were found by BEDTools' (version 2.26.0, Quinlan and Hall, 2010) sub-command 'intersect'. Pair-wise Pearson correlation coefficients between all samples were calculated based on the above mentioned first-base-in read coverages taking positions with a total sum of at least 10 reads in all samples combined into account. Positions that represent outliers with coverage values above the 99.99 percentile in one or more read libraries were not considered. The values were computed using the function 'corr' of the pandas Dataframe class (<https://doi.org/10.5281/zenodo.3509134>). For further details, please see the analysis scripts linked in the data and code availability section.

### RNA-seq analysis: Identification of OppZ targets

*V. cholerae* strains carrying either pBAD1K-ctrl or pBAD1K-*oppZ* were grown in biological triplicates to OD<sub>600</sub> of 0.5 and treated with 0.2% L-arabinose (final conc.). Cells were harvested after 15 min by addition of 0.2 volumes of stop mix (95% ethanol, 5% (v/v) phenol) and snap-frozen in liquid nitrogen. Total RNA was isolated and digested with TURBO DNase (Thermo Fisher Scientific). Ribosomal RNA was depleted using the Ribo-Zero kit for Gram-negative bacteria (#MRZGN126; Illumina, San Diego, California) and RNA integrity was confirmed with an Agilent 2100 Bioanalyzer. Directional cDNA libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (#E7760; NEB). The libraries were sequenced using a HiSeq 1500 System in single-read mode for 100 cycles. The read files in FASTQ format were imported into CLC Genomics Workbench v11 (RRID:SCR\_011853; Qiagen, Hilden, Germany) and trimmed for quality and 3' adaptors. Reads were mapped to the *V. cholerae* reference genome (NCBI accession numbers: NC\_002505.1 and

NC\_002506.1) including annotations for Vcr001-Vcr107 (*Papenfort et al., 2015b*) using the 'RNA-Seq Analysis' tool with standard parameters. Reads mapping in CDS were counted, and genes with a total count cut-off >15 in all samples were considered for analysis. Read counts were normalized (CPM), and transformed (log<sub>2</sub>). Differential expression was tested using the built-in tool corresponding to edgeR in exact mode with tagwise dispersions ('Empirical Analysis of DGE'). Genes with a fold change  $\geq 3.0$  and an FDR-adjusted p-value  $\leq 0.05$  were considered as differentially expressed.

### RNA-seq analysis: Bicyclomycin-dependent transcriptomes

*V. cholerae* oppA::3XFLAG oppB::3XFLAG oppF::3XFLAG strains with wild-type or mutated oppB start codon were grown in biological triplicates to OD<sub>600</sub> of 1.5, divided in half and treated with either bicyclomycin (25 µg/ml final conc.) or water. Cells were harvested after 120 min by addition of 0.2 volumes of stop mix (95% ethanol, 5% (v/v) phenol) and snap-frozen in liquid nitrogen. Total RNA was isolated and digested with TURBO DNase (Thermo Fisher Scientific). cDNA libraries were prepared by vertis Biotechnology AG in a 3' end-specific protocol: ribosomal RNA was depleted and the Illumina 5' sequencing adaptor was ligated to the 3' OH end of RNA molecules. First strand synthesis using M-MLV reverse transcriptase was followed by fragmentation and strand-specific ligation of the Illumina 3' sequencing adaptor to the 3' end of first-strand cDNA. Finally, 3' cDNA fragments were amplified, purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and sequenced using a NextSeq 500 system in single-read mode for 75 cycles. The read files in FASTQ format were imported into CLC Genomics Workbench v11 (Qiagen) and trimmed for quality and 3' adaptors. Reads were mapped to the *V. cholerae* reference genome (NCBI accession numbers: NC\_002505.1 and NC\_002506.1) including annotations for Vcr001-Vcr107 (*Papenfort et al., 2015b*) using the 'RNA-Seq Analysis' tool with standard parameters. Reads mapping in CDS were counted, and genes with a total count cut-off >8 in all samples were considered for analysis. Read counts were normalized (CPM), and transformed (log<sub>2</sub>). Differential expression was tested using the built in tool corresponding to edgeR in exact mode with tagwise dispersions ('Empirical Analysis of DGE'). Genes with a fold change  $\geq 3.0$  and an FDR-adjusted p-value  $\leq 0.05$  were considered as differentially expressed.

TIER-seq input data, analysis scripts and results are deposited at Zenodo (<https://doi.org/10.5281/zenodo.3750832>). Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author, Kai Papenfort ([kai.papenfort@uni-jena.de](mailto:kai.papenfort@uni-jena.de)).

### Acknowledgements

We thank Helmut Blum for help with the RNA sequencing experiments and Andreas Starick for excellent technical support. We thank Jörg Vogel, Gisela Storz, and Kathrin Fröhlich for comments on the manuscript and all members of the Papenfort lab for insightful discussions and suggestions.

### Additional information

#### Funding

Funder	Grant reference number	Author
Deutsche Forschungsgemeinschaft	GRK2062/1	Mona Hoyos Kai Papenfort
Vallee Foundation	Vallee Scholars program	Kai Papenfort
H2020 European Research Council	StG-758212	Michaela Huber Kai Papenfort
Human Frontier Science Program	CDA00024/2016-C	Mona Hoyos Michaela Huber Kai Papenfort
Deutsche Forschungsgemeinschaft	EXC 2051	Mona Hoyos Kai Papenfort
Deutsche Forschungsgemeinschaft	390713860	Mona Hoyos Kai Papenfort

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Mona Hoyos, Conceptualization, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - review and editing; Michaela Huber, Data curation, Validation, Investigation, Methodology, Writing - review and editing; Konrad U Förstner, Resources, Data curation, Software, Formal analysis, Writing - review and editing; Kai Papenfort, Conceptualization, Data curation, Supervision, Funding acquisition, Investigation, Methodology, Writing - original draft, Project administration, Writing - review and editing

### Author ORCIDs

Mona Hoyos  <https://orcid.org/0000-0003-1085-4723>

Kai Papenfort  <https://orcid.org/0000-0002-5560-9804>

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.58836.sa1>

Author response <https://doi.org/10.7554/eLife.58836.sa2>

## Additional files

### Supplementary files

- Supplementary file 1. TIER-seq sites in *Vibrio cholerae*.
- Supplementary file 2. RNase E-mediated maturation of sRNAs.
- Supplementary file 3. BCM-sensitive transcripts in *Vibrio cholerae*.
- Supplementary file 4. Bacterial strains, plasmids and DNA oligonucleotides.
- Transparent reporting form

### Data availability

All high-throughput sequencing data was deposited at GEO: GSE148675 (TIER-seq), GSE144479 (OppZ target identification) and GSE144478 (Term-Seq analysis). We have uploaded source data for all figures and figure supplements showing the numerical data from our TIER-seq analysis, the raw data from GFP and mKate fluorescence measurements, fold changes obtained from qRT-PCR experiments and fold changes obtained by the quantification of Western and Northern blots. Additionally, for all figures showing cropped images of Western or Northern blots, we show the full image and indicate the cropped area and the antibody or labelled oligonucleotide used to detect the signal.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Hoyos M, Huber M, Förstner K, Papenfort K	2020	Identification of bicyclomycin-sensitive transcripts in <i>Vibrio cholerae</i>	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144478">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144478</a>	NCBI Gene Expression Omnibus, GSE144478
Hoyos M, Huber M, Förstner KU, Papenfort K	2020	Global identification of RNase E sites in <i>Vibrio cholerae</i>	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE148675">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE148675</a>	NCBI Gene Expression Omnibus, GSE148675
Hoyos M, Huber M, Förstner K, Papenfort K	2020	OppZ target identification	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144479">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144479</a>	NCBI Gene Expression Omnibus, GSE144479

## References

- Adams PP, Storz G. 2020. Prevalence of small base-pairing RNAs derived from diverse genomic loci. *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1863**:194524. DOI: <https://doi.org/10.1016/j.bbagr.2020.194524>
- Afroz T, Beisel CL. 2013. Understanding and exploiting feedback in synthetic biology. *Chemical Engineering Science* **103**:79–90. DOI: <https://doi.org/10.1016/j.ces.2013.02.017>
- Ait-Bara S, Carpousis AJ. 2015. RNA degradosomes in Bacteria and chloroplasts: classification, distribution and evolution of RNase E homologs. *Molecular Microbiology* **97**:1021–1135. DOI: <https://doi.org/10.1111/mmi.13095>, PMID: 26096689
- Allfano P, Rivellini F, Limauro D, Bruni CB, Carlomagno MS. 1991. A consensus motif common to all rho-dependent prokaryotic transcription terminators. *Cell* **64**:553–563. DOI: [https://doi.org/10.1016/0092-8674\(91\)90239-U](https://doi.org/10.1016/0092-8674(91)90239-U)
- Alon U. 2007. Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8**:450–461. DOI: <https://doi.org/10.1038/nrg2102>, PMID: 17510665
- Apirion D, Lassar AB. 1978. A conditional lethal mutant of *Escherichia coli* which affects the processing of ribosomal RNA. *The Journal of Biological Chemistry* **253**:1738–1742. PMID: 342528
- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the keio collection. *Molecular Systems Biology* **2**:0008. DOI: <https://doi.org/10.1038/msb4100050>, PMID: 16738554
- Bandyra KJ, Luisi BF. 2018. RNase E and the High-Fidelity orchestration of RNA metabolism. *Microbiology Spectrum* **6**:ch1. DOI: <https://doi.org/10.1128/9781683670247.ch1>
- Bechhofer DH, Deutscher MP. 2019. Bacterial ribonucleases and their roles in RNA metabolism. *Critical Reviews in Biochemistry and Molecular Biology* **54**:242–300. DOI: <https://doi.org/10.1080/10409238.2019.1651816>, PMID: 31464530
- 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 & Development* **26**:1864–1873. DOI: <https://doi.org/10.1101/gad.195412.112>, PMID: 22895254
- Bossi L, Figueroa-Bossi N, Bouloc P, Boudvillain M. 2020. Regulatory interplay between small RNAs and transcription termination factor rho. *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1863**:194546. DOI: <https://doi.org/10.1016/j.bbagr.2020.194546>
- Brosse A, Korobeinikova A, Gottesman S, Guillier M. 2016. Unexpected properties of sRNA promoters allow feedback control via regulation of a two-component system. *Nucleic Acids Research* **177**:642–9666. DOI: <https://doi.org/10.1093/nar/gkw642>
- Cameron DE, Urbach JM, Mekalanos JJ. 2008. A defined transposon mutant library and its use in identifying motility genes in *Vibrio cholerae*. *PNAS* **105**:8736–8741. DOI: <https://doi.org/10.1073/pnas.0803281105>, PMID: 18574146
- Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E. 2008. Termination factor rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* **320**:935–938. DOI: <https://doi.org/10.1126/science.1152763>, PMID: 18487194
- Castellana M, Wilson MZ, Xu Y, Joshi P, Cristea IM, Rabinowitz JD, Gitai Z, Wingreen NS. 2014. Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nature Biotechnology* **32**:1011–1018. DOI: <https://doi.org/10.1038/nbt.3018>, PMID: 25262299
- Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. 2012. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *The EMBO Journal* **31**:4005–4019. DOI: <https://doi.org/10.1038/emboj.2012.229>, PMID: 22922465
- Chao Y, Li L, Girodat D, Förstner KU, Said N, Corcoran C, Śmiga M, Papenfort K, Reinhardt R, Wieden HJ, Luisi BF, Vogel J. 2017. In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Molecular Cell* **65**:39–51. DOI: <https://doi.org/10.1016/j.molcel.2016.11.002>, PMID: 28061332
- Chao Y, Vogel J. 2016. A 3' UTR-Derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. *Molecular Cell* **61**:352–363. DOI: <https://doi.org/10.1016/j.molcel.2015.12.023>, PMID: 26805574
- Charlier D, Nguyen Le Minh P, Roovers M. 2018. Regulation of carbamoylphosphate synthesis in *Escherichia coli*: an amazing metabolite at the crossroad of arginine and pyrimidine biosynthesis. *Amino Acids* **50**:1647–1661. DOI: <https://doi.org/10.1007/s00726-018-2654-z>, PMID: 30238253
- Ciampi MS. 2006. Rho-dependent terminators and transcription termination. *Microbiology* **152**:2515–2528. DOI: <https://doi.org/10.1099/mic.0.28982-0>, PMID: 16946247
- Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC, Vogel J. 2012. Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Molecular Microbiology* **84**:428–445. DOI: <https://doi.org/10.1111/j.1365-2958.2012.08031.x>, PMID: 22458297
- Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16**:10881–10890. DOI: <https://doi.org/10.1093/nar/16.22.10881>, PMID: 2849754
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Research* **14**:1188–1190. DOI: <https://doi.org/10.1101/gr.849004>, PMID: 15173120
- Dar D, Shamir M, Mellin JR, Koutero M, Stern-Ginossar N, Cossart P, Sorek R. 2016. Term-seq reveals abundant ribo-regulation of antibiotics resistance in Bacteria. *Science* **352**:aad9822. DOI: <https://doi.org/10.1126/science.aad9822>, PMID: 27120414

- Dar D, Sorek R. 2018a. Bacterial noncoding RNAs excised from within Protein-Coding transcripts. *mBio* **9**:01730–18. DOI: <https://doi.org/10.1128/mBio.01730-18>
- Dar D, Sorek R. 2018b. High-resolution RNA 3'-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Research* **46**:6797–6805. DOI: <https://doi.org/10.1093/nar/gky274>, PMID: 29669055
- Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS* **97**:6640–6645. DOI: <https://doi.org/10.1073/pnas.120163297>, PMID: 10829079
- Davis BM, Waldor MK. 2007. RNase E-dependent processing stabilizes MicX, a *Vibrio cholerae* sRNA. *Molecular Microbiology* **65**:373–385. DOI: <https://doi.org/10.1111/j.1365-2958.2007.05796.x>, PMID: 17590231
- De Mets F, Van Melderen L, Gottesman S. 2019. Regulation of acetate metabolism and coordination with the TCA cycle via a processed small RNA. *PNAS* **116**:1043–1052. DOI: <https://doi.org/10.1073/pnas.1815288116>, PMID: 30591570
- Deana A, Celesnik H, Belasco JG. 2008. The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* **451**:355–358. DOI: <https://doi.org/10.1038/nature06475>, PMID: 18202662
- Doeven MK, Abele R, Tampé R, Poolman B. 2004. The binding specificity of OppA determines the selectivity of the oligopeptide ATP-binding cassette transporter. *Journal of Biological Chemistry* **279**:32301–32307. DOI: <https://doi.org/10.1074/jbc.M404343200>, PMID: 15169767
- Dunn AK, Millikan DS, Adin DM, Bose JL, Stabb EV. 2006. New *rfp*- and *pES213*-derived tools for analyzing symbiotic *Vibrio fischeri* reveal patterns of infection and *lux* expression in situ. *Applied and Environmental Microbiology* **72**:802–810. DOI: <https://doi.org/10.1128/AEM.72.1.802-810.2006>, PMID: 16391121
- Feng L, Rutherford ST, Papenfort K, Bagert JD, van Kessel JC, Tirrell DA, Wingreen NS, Bassler BL. 2015. A *qrr* noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell* **160**:228–240. DOI: <https://doi.org/10.1016/j.cell.2014.11.051>, PMID: 25579683
- Förstner KU, Vogel J, Sharma CM. 2014. READemption—a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics* **30**:3421–3423. DOI: <https://doi.org/10.1093/bioinformatics/btu533>, PMID: 25123900
- Förstner KU, Reuscher CM, Haberzettl K, Weber L, Klug G. 2018. RNase E cleavage shapes the transcriptome of *Rhodobacter sphaeroides* and strongly impacts phototrophic growth. *Life Science Alliance* **1**:e201800080. DOI: <https://doi.org/10.26508/lsa.201800080>, PMID: 30456366
- Fröhlich KS, Papenfort K, Fekete A, Vogel J. 2013. A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *The EMBO Journal* **32**:2963–2979. DOI: <https://doi.org/10.1038/emboj.2013.222>, PMID: 24141880
- Gorski SA, Vogel J, Doudna JA. 2017. RNA-based recognition and targeting: sowing the seeds of specificity. *Nature Reviews Molecular Cell Biology* **18**:215–228. DOI: <https://doi.org/10.1038/nrm.2016.174>, PMID: 28196981
- Grabowicz M, Koren D, Silhavy TJ. 2016. The CpxQ sRNA negatively regulates *skp* to prevent mistargeting of  $\beta$ -Barrel outer membrane proteins into the cytoplasmic membrane. *mBio* **7**:00312–16. DOI: <https://doi.org/10.1128/mBio.00312-16>
- Gruber AJ, Zavolan M. 2019. Alternative cleavage and polyadenylation in health and disease. *Nature Reviews Genetics* **20**:599–614. DOI: <https://doi.org/10.1038/s41576-019-0145-z>, PMID: 31267064
- 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**:6781–6794. DOI: <https://doi.org/10.1093/nar/gkn742>, PMID: 18953042
- Herzog R, Peschek N, Fröhlich KS, Schumacher K, Papenfort K. 2019. Three autoinducer molecules act in concert to control virulence gene expression in *Vibrio cholerae*. *Nucleic Acids Research* **47**:3171–3183. DOI: <https://doi.org/10.1093/nar/gky1320>, PMID: 30649554
- Hiles ID, Gallagher MP, Jamieson DJ, Higgins CF. 1987. Molecular characterization of the oligopeptide permease of *Salmonella typhimurium*. *Journal of Molecular Biology* **195**:125–142. DOI: [https://doi.org/10.1016/0022-2836\(87\)90332-9](https://doi.org/10.1016/0022-2836(87)90332-9), PMID: 2821267
- Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermüller J. 2009. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLOS Computational Biology* **5**:e1000502. DOI: <https://doi.org/10.1371/journal.pcbi.1000502>, PMID: 19750212
- Holmqvist E, Vogel J. 2018. RNA-binding proteins in bacteria. *Nature Reviews Microbiology* **16**:601–615. DOI: <https://doi.org/10.1038/s41579-018-0049-5>, PMID: 29995832
- Hör J, Gorski SA, Vogel J. 2018. Bacterial RNA Biology on a Genome Scale. *Molecular Cell* **70**:785–799. DOI: <https://doi.org/10.1016/j.molcel.2017.12.023>
- Huber M, Fröhlich KS, Radmer J, Papenfort K. 2020. Switching fatty acid metabolism by an RNA-controlled feed forward loop. *PNAS* **117**:8044–8054. DOI: <https://doi.org/10.1073/pnas.1920753117>, PMID: 32193348
- Hussein R, Lim HN. 2012. Direct comparison of small RNA and transcription factor signaling. *Nucleic Acids Research* **40**:7269–7279. DOI: <https://doi.org/10.1093/nar/gks439>, PMID: 22618873
- Kavita K, de Mets F, Gottesman S. 2018. New aspects of RNA-based regulation by *hfq* and its partner sRNAs. *Current Opinion in Microbiology* **42**:53–61. DOI: <https://doi.org/10.1016/j.mib.2017.10.014>, PMID: 29125938
- Kelly CL, Harris AWK, Steel H, Hancock EJ, Heap JT, Papachristodoulou A. 2018. Synthetic negative feedback circuits using engineered small RNAs. *Nucleic Acids Research* **46**:9875–9889. DOI: <https://doi.org/10.1093/nar/gky828>, PMID: 30212900
- Kimura S, Waldor MK. 2019. The RNA degradosome promotes tRNA quality control through clearance of hypomodified tRNA. *PNAS* **116**:1394–1403. DOI: <https://doi.org/10.1073/pnas.1814130116>, PMID: 30622183

- Korkmaz G**, Holm M, Wiens T, Sanyal S. 2014. Comprehensive analysis of stop Codon usage in Bacteria and its correlation with release factor abundance. *Journal of Biological Chemistry* **289**:30334–30342. DOI: <https://doi.org/10.1074/jbc.M114.606632>, PMID: 25217634
- Lin P**, Pu Q, Wu Q, Zhou C, Wang B, Schettler J, Wang Z, Qin S, Gao P, Li R, Li G, Cheng Z, Lan L, Jiang J, Wu M. 2019. High-throughput screen reveals sRNAs regulating crRNA biogenesis by targeting CRISPR leader to repress rho termination. *Nature Communications* **10**:3728. DOI: <https://doi.org/10.1038/s41467-019-11695-8>, PMID: 31427601
- Lorenz R**, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA package 2.0. *Algorithms for Molecular Biology* **6**:26. DOI: <https://doi.org/10.1186/1748-7188-6-26>, PMID: 22115189
- Love MI**, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550. DOI: <https://doi.org/10.1186/s13059-014-0550-8>, PMID: 25516281
- Mackie GA**. 2013. RNase E: at the interface of bacterial RNA processing and decay. *Nature Reviews Microbiology* **11**:45–57. DOI: <https://doi.org/10.1038/nrmicro2930>, PMID: 23241849
- Malka Y**, Steiman-Shimony A, Rosenthal E, Argaman L, Cohen-Daniel L, Arbib E, Margalit H, Kaplan T, Berger M. 2017. Post-transcriptional 3'-UTR cleavage of mRNA transcripts generates thousands of stable uncapped autonomous RNA fragments. *Nature Communications* **8**:2029. DOI: <https://doi.org/10.1038/s41467-017-02099-7>, PMID: 29229900
- Martin M**. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**:10–12. DOI: <https://doi.org/10.14806/ej.17.1.200>
- Massé E**, Escorcía FE, Gottesman S. 2003. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes & Development* **17**:2374–2383. DOI: <https://doi.org/10.1101/gad.1127103>, PMID: 12975324
- Melamed S**, Peer A, Faigenbaum-Romm R, Gatt YE, Reiss N, Bar A, Altuvia Y, Argaman L, Margalit H. 2016. Global mapping of small RNA-Target interactions in Bacteria. *Molecular Cell* **63**:884–897. DOI: <https://doi.org/10.1016/j.molcel.2016.07.026>, PMID: 27588604
- Melamed S**, Adams PP, Zhang A, Zhang H, Storz G. 2020. RNA-RNA interactomes of ProQ and hfq reveal overlapping and competing roles. *Molecular Cell* **77**:411–425. DOI: <https://doi.org/10.1016/j.molcel.2019.10.022>
- Mitra P**, Ghosh G, Hafeezunnisa M, Sen R. 2017. Rho protein: roles and mechanisms. *Annual Review of Microbiology* **71**:687–709. DOI: <https://doi.org/10.1146/annurev-micro-030117-020432>, PMID: 28731845
- Miyakoshi M**, Chao Y, Vogel J. 2015. Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Current Opinion in Microbiology* **24**:132–139. DOI: <https://doi.org/10.1016/j.mib.2015.01.013>, PMID: 25677420
- Miyakoshi M**, Matera G, Maki K, Sone Y, Vogel J. 2019. Functional expansion of a TCA cycle operon mRNA by a 3' end-derived small RNA. *Nucleic Acids Research* **47**:2075–2088. DOI: <https://doi.org/10.1093/nar/gky1243>, PMID: 30541135
- Mohanty BK**, Kushner SR. 2018. Enzymes involved in posttranscriptional RNA metabolism in Gram-Negative Bacteria. *Microbiology Spectrum* **6**:2017. DOI: <https://doi.org/10.1128/microbiolspec.RWR-0011-2017>
- Morita T**, Maki K, Aiba H. 2005. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes & Development* **19**:2176–2186. DOI: <https://doi.org/10.1101/gad.1330405>, PMID: 16166379
- Nadiras C**, Eveno E, Schwartz A, Figueroa-Bossi N, Boudvillain M. 2018. A multivariate prediction model for Rho-dependent termination of transcription. *Nucleic Acids Research* **46**:8245–8260. DOI: <https://doi.org/10.1093/nar/gky563>, PMID: 29931073
- Nitzan M**, Rehani R, Margalit H. 2017. Integration of bacterial small RNAs in regulatory networks. *Annual Review of Biophysics* **46**:131–148. DOI: <https://doi.org/10.1146/annurev-biophys-070816-034058>, PMID: 28532217
- Papenfort K**, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J. 2013. Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* **153**:426–437. DOI: <https://doi.org/10.1016/j.cell.2013.03.003>, PMID: 23582330
- Papenfort K**, Espinosa E, Casadesús J, Vogel J. 2015a. Small RNA-based feedforward loop with AND-gate logic regulates extrachromosomal DNA transfer in *Salmonella*. *PNAS* **112**:E4772–E4781. DOI: <https://doi.org/10.1073/pnas.1507825112>, PMID: 26307765
- Papenfort K**, Förstner KU, Cong JP, Sharma CM, Bassler BL. 2015b. Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *PNAS* **112**:E766–E775. DOI: <https://doi.org/10.1073/pnas.1500203112>, PMID: 25646441
- Papenfort K**, Silpe JE, Schramma KR, Cong JP, Seyedsayamdost MR, Bassler BL. 2017. A *Vibrio cholerae* autoinducer-receptor pair that controls biofilm formation. *Nature Chemical Biology* **13**:551–557. DOI: <https://doi.org/10.1038/nchembio.2336>, PMID: 28319101
- Peschek N**, Hoyos M, Herzog R, Förstner KU, Papenfort K. 2019. A conserved RNA seed-pairing domain directs small RNA-mediated stress resistance in enterobacteria. *The EMBO Journal* **38**:e101650. DOI: <https://doi.org/10.15252/embj.2019101650>, PMID: 31313835
- Peters JM**, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R. 2012. Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes & Development* **26**:2621–2633. DOI: <https://doi.org/10.1101/gad.196741.112>, PMID: 23207917
- Pu M**, Chen J, Tao Z, Miao L, Qi X, Wang Y, Ren J. 2019. Regulatory network of miRNA on its target: coordination between transcriptional and post-transcriptional regulation of gene expression. *Cellular and Molecular Life Sciences* **76**:441–451. DOI: <https://doi.org/10.1007/s00018-018-2940-7>, PMID: 30374521

- Qi LS, Arkin AP. 2014. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nature Reviews. Microbiology* **12**:341–354. DOI: <https://doi.org/10.1038/nrmicro3244>, PMID: 24736794
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. DOI: <https://doi.org/10.1093/bioinformatics/btq033>, PMID: 20110278
- Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. 2004. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**:1507–1517. DOI: <https://doi.org/10.1261/rna.5248604>, PMID: 15383676
- Richards J, Belasco JG. 2019. Obstacles to scanning by RNase E govern bacterial mRNA lifetimes by hindering access to distal cleavage sites. *Molecular Cell* **74**:284–295. DOI: <https://doi.org/10.1016/j.molcel.2019.01.044>
- Rosenfeld N, Elowitz MB, Alon U. 2002. Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology* **323**:785–793. DOI: [https://doi.org/10.1016/S0022-2836\(02\)00994-4](https://doi.org/10.1016/S0022-2836(02)00994-4), PMID: 12417193
- Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH image to ImageJ: 25 years of image analysis. *Nature Methods* **9**:671–675. DOI: <https://doi.org/10.1038/nmeth.2089>, PMID: 22930834
- Sedyarova N, Shamovsky I, Bharati BK, Epshtein V, Chen J, Gottesman S, Schroeder R, Nudler E. 2016. sRNA-Mediated control of transcription termination in *E. coli*. *Cell* **167**:111–121. DOI: <https://doi.org/10.1016/j.cell.2016.09.004>
- Shen-Orr SS, Milo R, Mangan S, Alon U. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* **31**:64–68. DOI: <https://doi.org/10.1038/ng881>, PMID: 11967538
- Simon R, Prierer U, Pühler A. 1983. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative Bacteria. *Bio/Technology* **1**:784–791. DOI: <https://doi.org/10.1038/nbt1183-784>
- Skorupski K, Taylor RK. 1996. Positive selection vectors for allelic exchange. *Gene* **169**:47–52. DOI: [https://doi.org/10.1016/0378-1119\(95\)00793-8](https://doi.org/10.1016/0378-1119(95)00793-8), PMID: 8635748
- Soper T, Mandin P, Majdalani N, Gottesman S, Woodson SA. 2010. Positive regulation by small RNAs and the role of hfq. *PNAS* **107**:9602–9607. DOI: <https://doi.org/10.1073/pnas.1004435107>, PMID: 20457943
- Svenningsen SL, Tu KC, Bassler BL. 2009. Gene dosage compensation calibrates four regulatory RNAs to control *Vibrio cholerae* quorum sensing. *The EMBO Journal* **28**:429–439. DOI: <https://doi.org/10.1038/emboj.2008.300>, PMID: 19165149
- Thelin KH, Taylor RK. 1996. Toxin-coregulated pilus, but not mannose-sensitive hemagglutinin, is required for colonization by *Vibrio cholerae* O1 el Tor biotype and O139 strains. *Infection and Immunity* **64**:2853–2856. DOI: <https://doi.org/10.1128/IAI.64.7.2853-2856.1996>, PMID: 8698524
- Updegrove TB, Kouse AB, Bandyra KJ, Storz G. 2019. Stem-loops direct precise processing of 3' UTR-derived small RNA MicL. *Nucleic Acids Research* **47**:1482–1492. DOI: <https://doi.org/10.1093/nar/gky1175>
- Urban JH, Vogel J. 2007. Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Research* **35**:1018–1037. DOI: <https://doi.org/10.1093/nar/gkl1040>, PMID: 17264113
- Wang X, Ji SC, Jeon HJ, Lee Y, Lim HM. 2015. Two-level inhibition of *galK* expression by spot 42: degradation of mRNA mK2 and enhanced transcription termination before the *galK* gene. *PNAS* **112**:7581–7586. DOI: <https://doi.org/10.1073/pnas.1424683112>, PMID: 26045496
- Wang C, Chao Y, Matera G, Gao Q, Vogel J. 2020. The conserved 3' UTR-derived small RNA NarS mediates mRNA crossregulation during nitrate respiration. *Nucleic Acids Research* **48**:2126–2143. DOI: <https://doi.org/10.1093/nar/gkz1168>, PMID: 31863581
- Waters CM, Bassler BL. 2006. The *Vibrio harveyi* quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers. *Genes & Development* **20**:2754–2767. DOI: <https://doi.org/10.1101/gad.1466506>, PMID: 17015436
- Zwiefka A, Kohn H, Widger WR. 1993. Transcription termination factor rho: the site of bicyclomycin inhibition in *Escherichia coli*. *Biochemistry* **32**:3564–3570. DOI: <https://doi.org/10.1021/bi00065a007>, PMID: 8466900

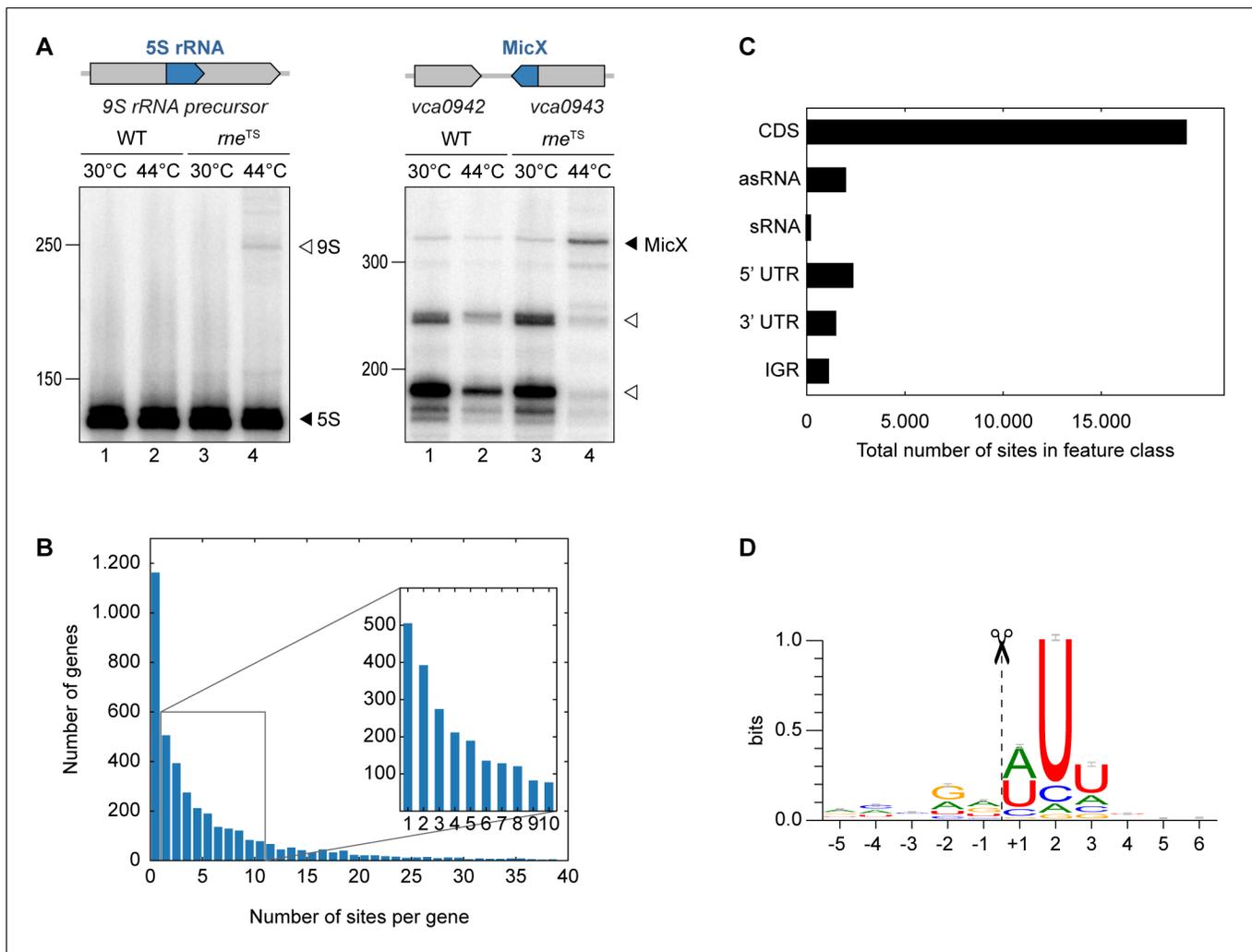


---

## Figures and figure supplements

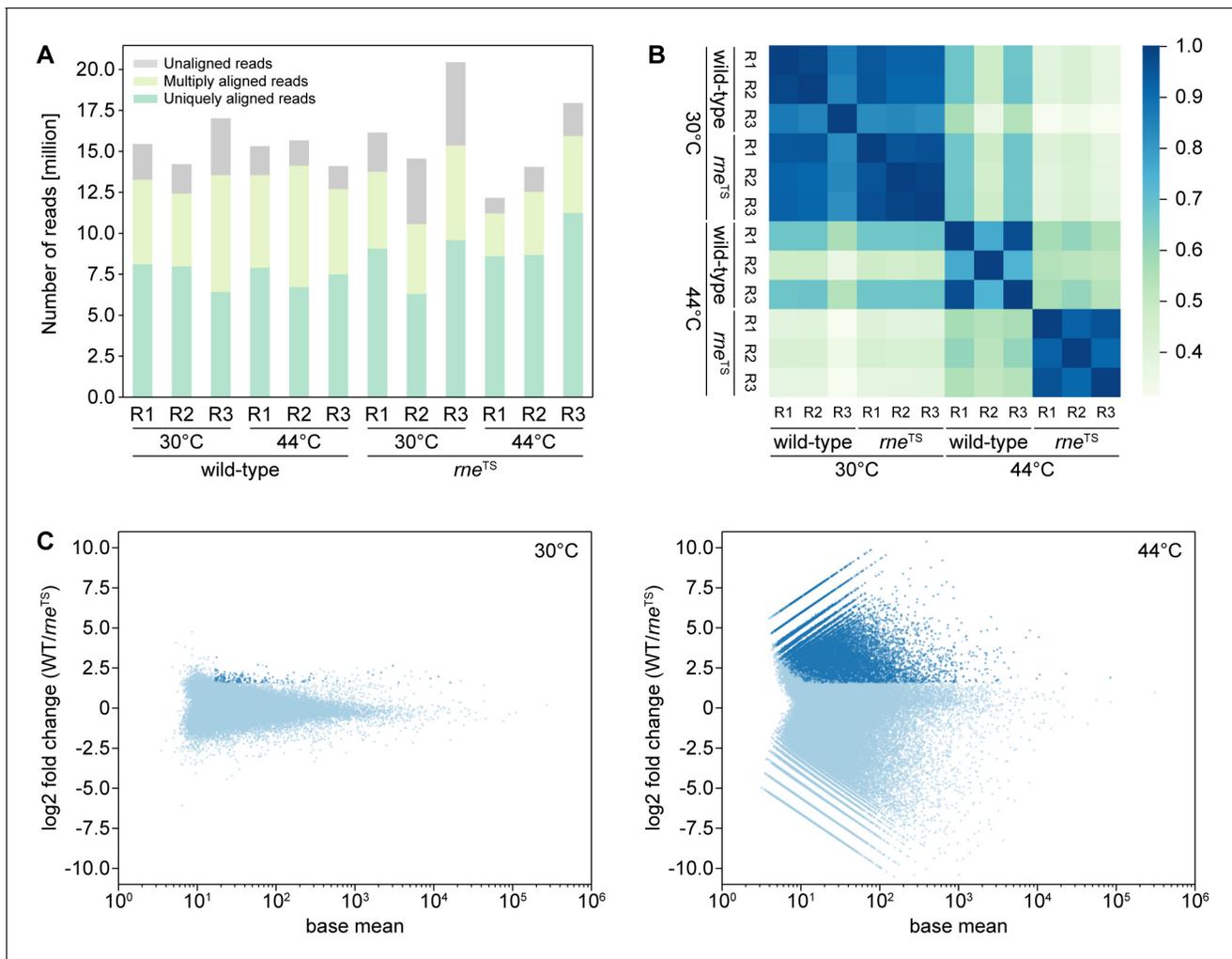
Gene autoregulation by 3' UTR-derived bacterial small RNAs

**Mona Hoyos *et al***

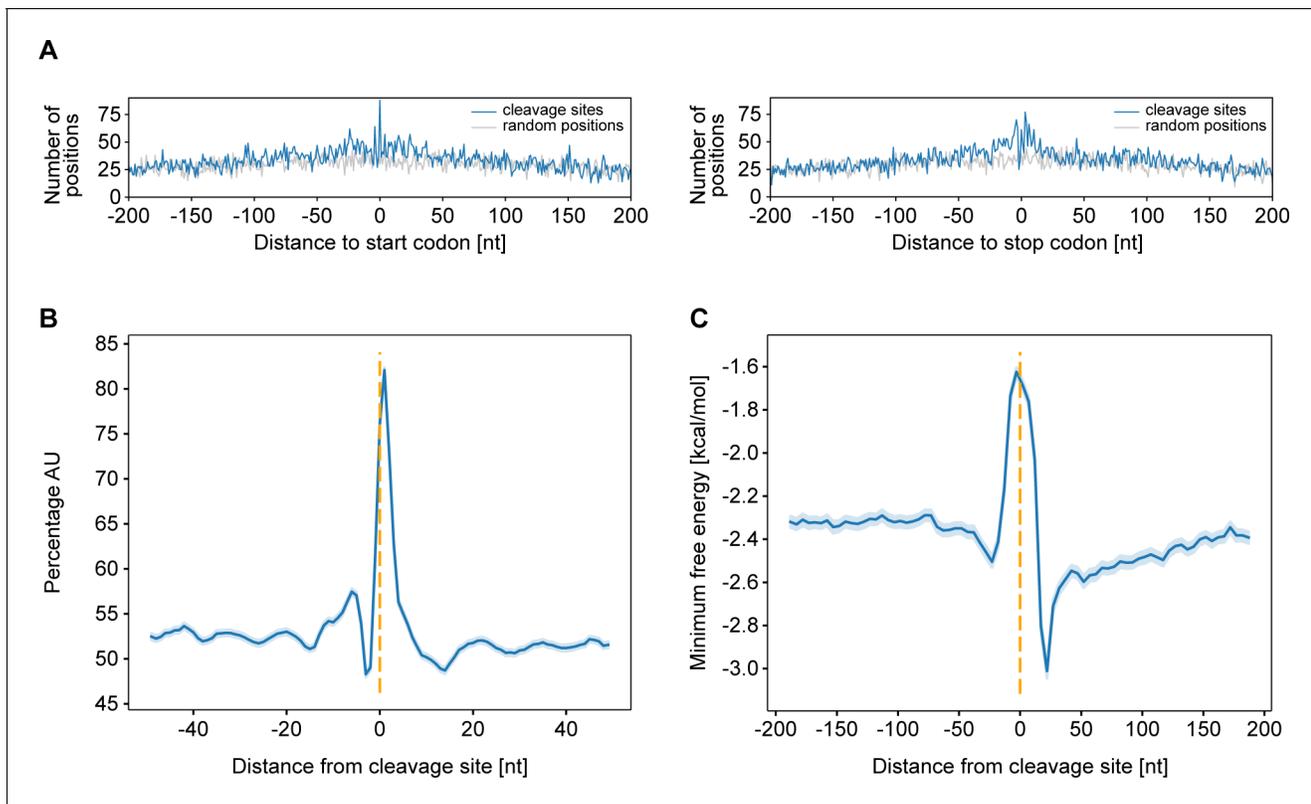


**Figure 1.** TIER-seq analysis of *V. cholerae*. (A) *V. cholerae* wild-type and *rne<sup>TS</sup>* strains were grown at 30°C to stationary phase (OD<sub>600</sub> of 2.0). Cultures were divided in half and continuously grown at either 30°C or 44°C for 60 min. Cleavage patterns of 5S rRNA and 3' UTR-derived MicX were analyzed on Northern blots. Closed triangles indicate mature 5S or full-length MicX, open triangles indicate the 9S precursor or MicX processing products. (B, C, D) Biological triplicates of *V. cholerae* wild-type and *rne<sup>TS</sup>* strains were grown at 30°C to late exponential phase (OD<sub>600</sub> of 1.0). Cultures were divided in half and continuously grown at either 30°C or 44°C for 60 min. Isolated RNA was subjected to RNA-seq and RNase E cleavage sites were determined as described in the materials and methods section. (B) Number of cleavage sites detected per gene. (C) Classification of RNase E sites by their genomic location. (D) The RNase E consensus motif based on all detected cleavage sites. The total height of the error bar is twice the small sample correction.

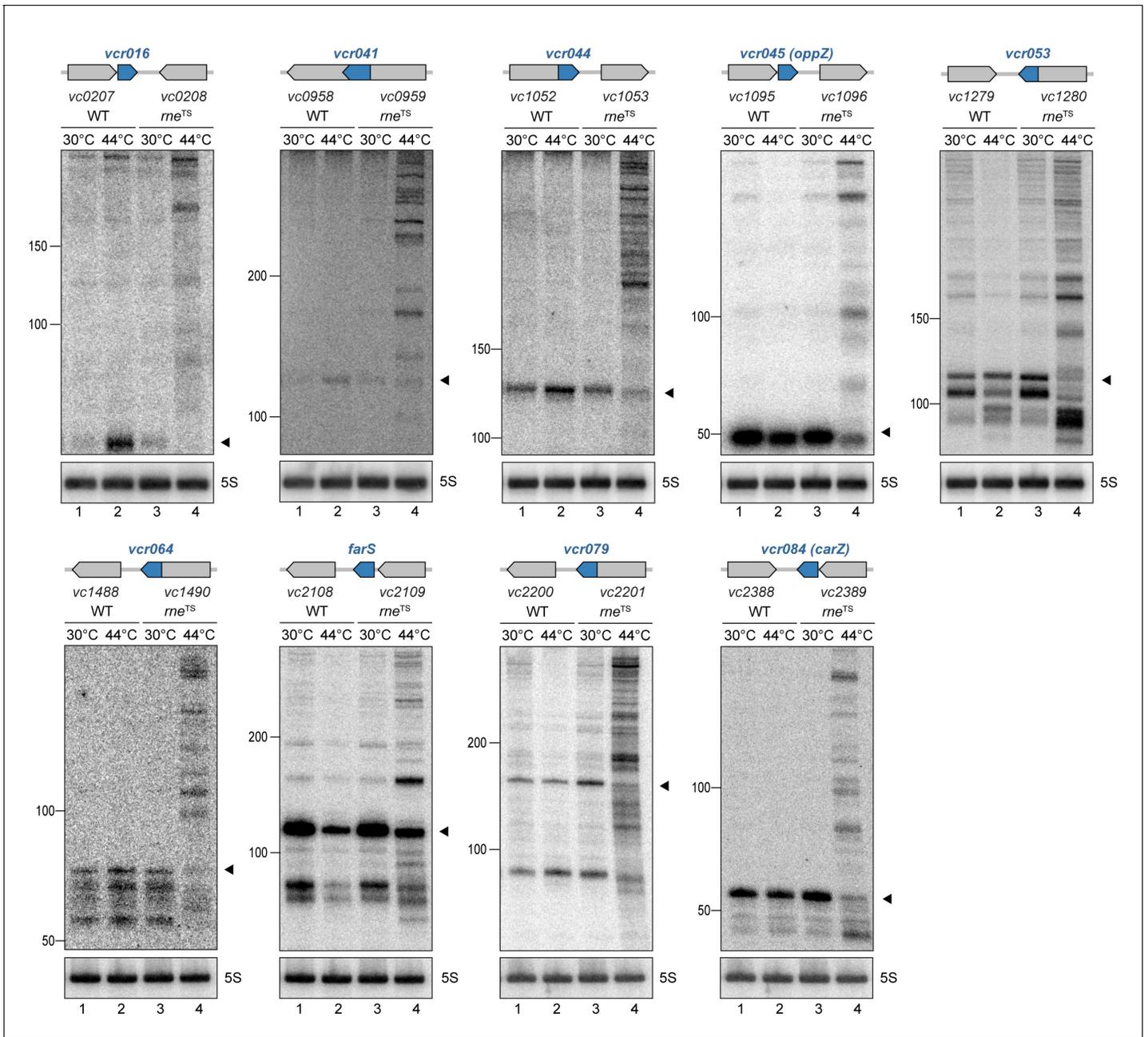




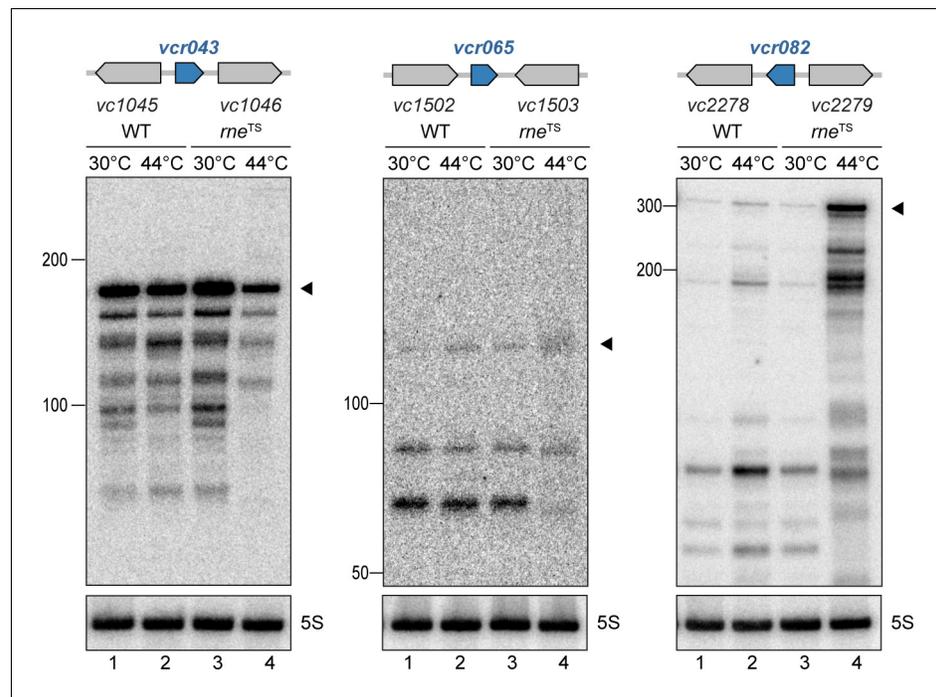
**Figure 1—figure supplement 2.** TIER-Seq read mapping statistics. TIER-seq was performed as described in **Figure 1**. **(A)** Total number of raw cDNA reads obtained for all samples, showing the fractions of uniquely aligned reads (dark green), multiply aligned reads (light green) or unaligned reads (grey). R1-R3 indicate the biological triplicates. **(B)** Similarity of 5' ends profiles of uniquely aligned reads, obtained by comparison of all detected 5' end positions between the respective cDNA libraries. Colored rectangles show the Pearson correlation coefficient corresponding to the scale bar on the right. **(C)** Global analysis of 5' profiles at the permissive (30°C, left) and non-permissive temperature (44°C, right). Plots show average coverage levels of 5' read ends and the respective log<sub>2</sub> fold change in wild-type samples compared to *rne<sup>TS</sup>* samples. Candidate RNase E cleavage sites were determined as positions enriched  $\geq 3$  fold in the wild-type ( $p$ -value $<0.05$ ) and are shown in dark blue.



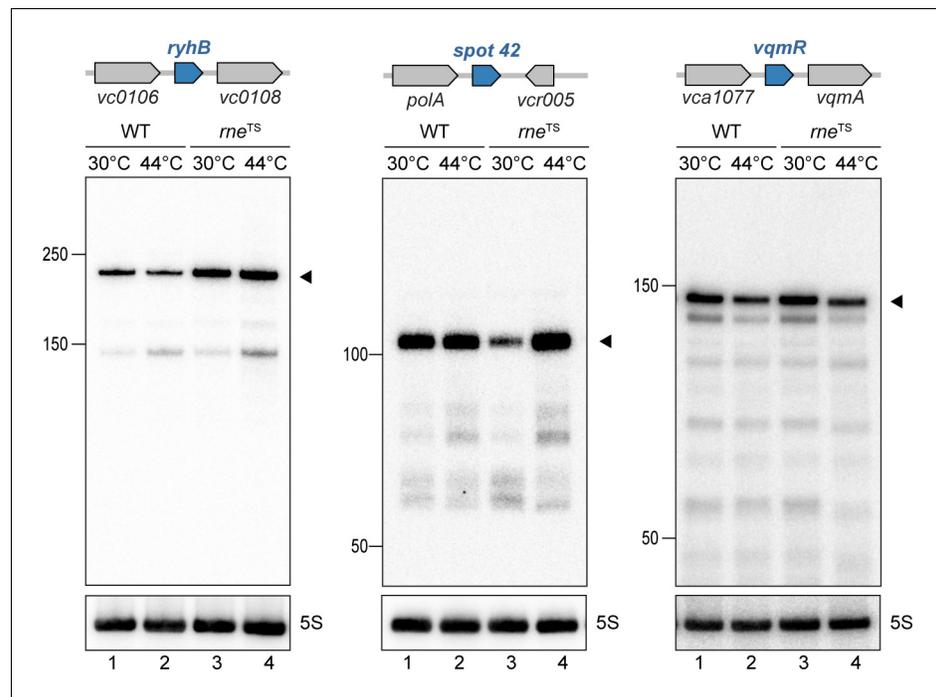
**Figure 1—figure supplement 3.** Position and characteristics of RNase E cleavage sites. TIER-seq was performed as described in **Figure 1**. (A) Frequency of RNase E sites or the same number of randomly selected genome positions dependent on their relative position to start codons (left) and stop codons (right). (B) AU content around the RNase E cleavage sites. The 95% confidence interval is indicated in light blue. (C) Degree of RNA structure around RNase E cleavage sites. Minimal folding energy (MFE) was calculated in five nt steps for each 25 nt window. The 95% confidence interval is indicated in light blue.



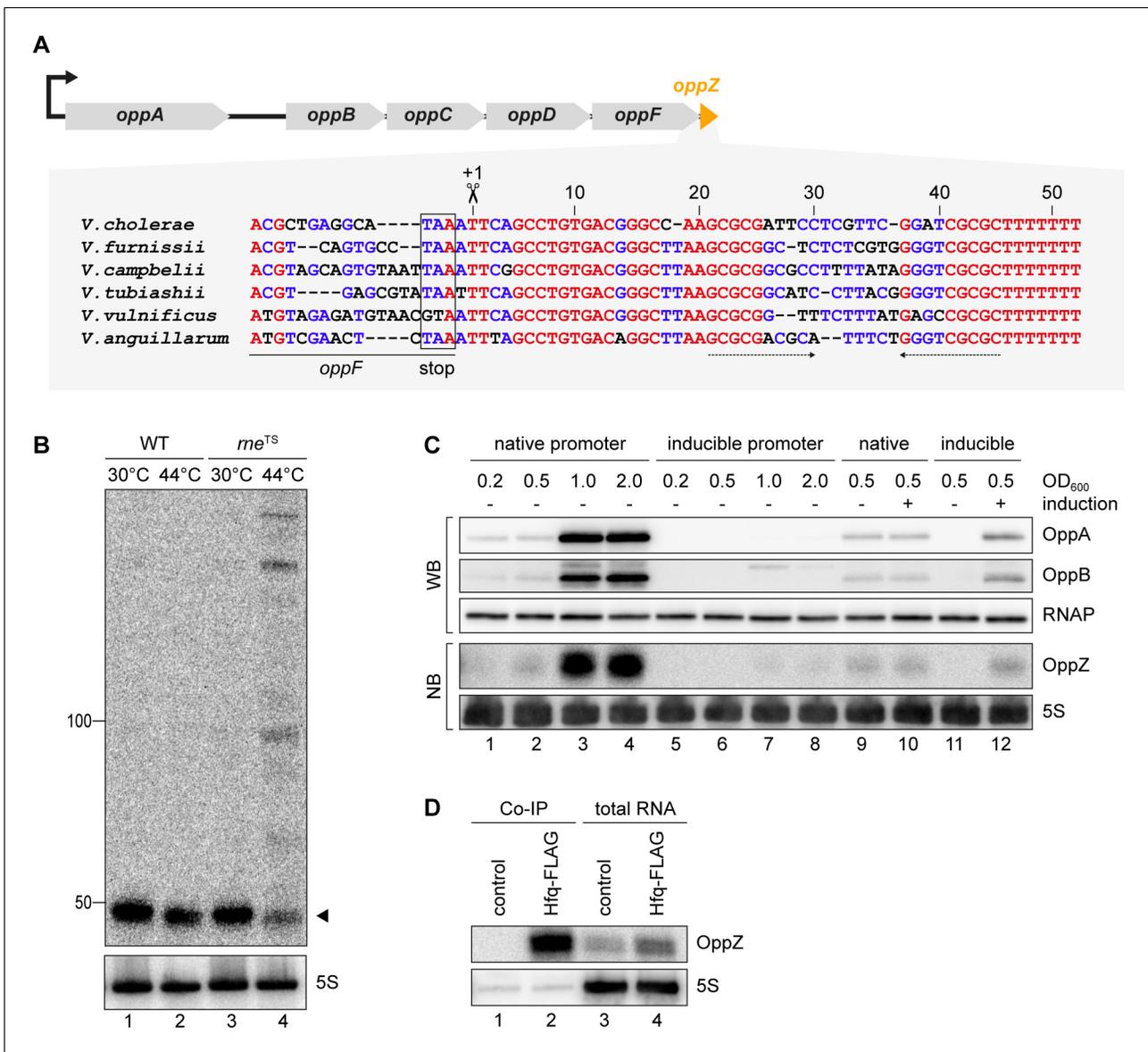
**Figure 1—figure supplement 4.** RNase E-mediated maturation of sRNAs from 3' UTRs. *V. cholerae* wild-type and *me<sup>TS</sup>* strains were grown at 30°C to stationary phase (OD<sub>600</sub> of 2.0). Cultures were divided in half and continuously grown at either 30°C or 44°C for 30 min. Cleavage patterns of 3' UTR-derived sRNAs were monitored on Northern blots. The genomic locations and relative orientations are shown above the gels. Genes are shown in gray; sRNAs are shown in blue. Filled triangles indicate the size of mature sRNAs. 5S rRNA served as loading control.



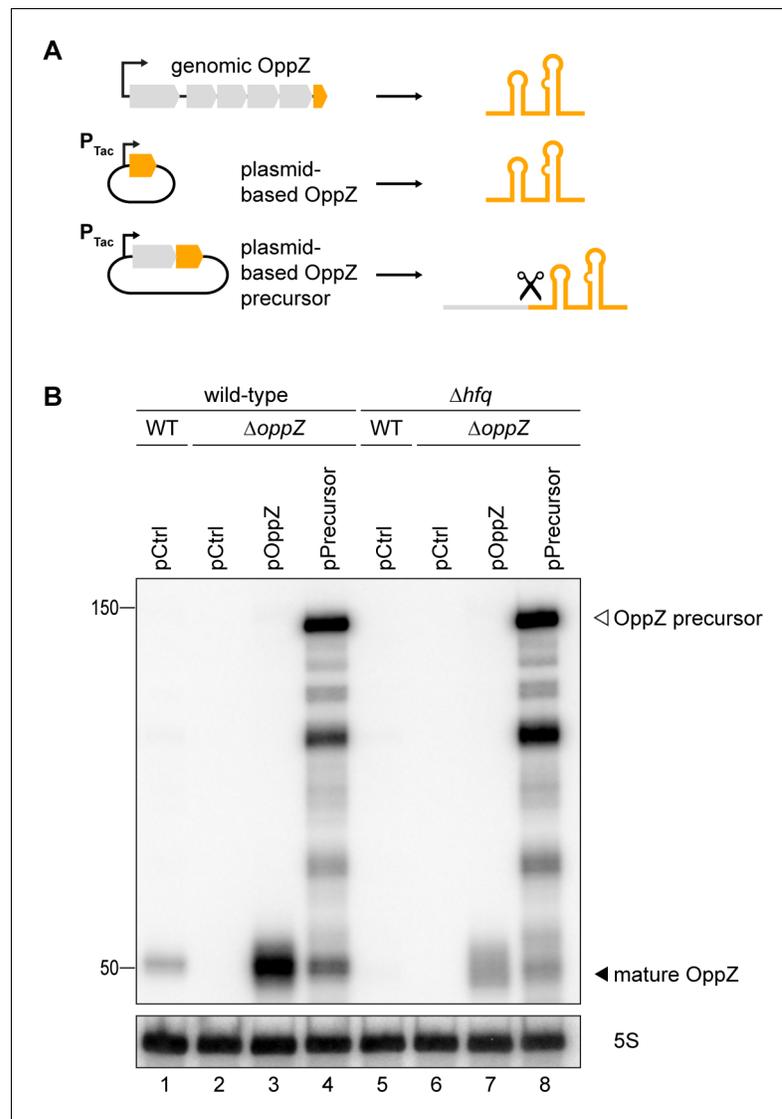
**Figure 1—figure supplement 5.** RNase E-mediated maturation of sRNAs from IGRs. *V. cholerae* wild-type and *rne<sup>TS</sup>* strains were grown at 30°C to stationary phase ( $OD_{600}$  of 2.0). Cultures were divided in half and continuously grown at either 30°C or 44°C for 30 min. Cleavage patterns of intergenic sRNAs were monitored on Northern blots. The genomic locations and relative orientations are shown above the gels. Genes are shown in gray; sRNAs are shown in blue. Filled triangles indicate the size of unprocessed sRNAs. 5S rRNA served as loading control.



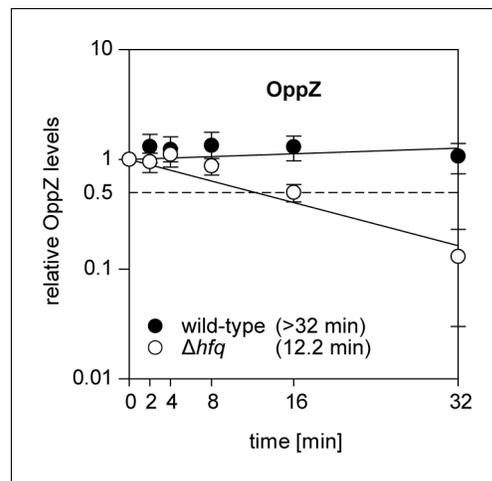
**Figure 1—figure supplement 6.** Expression of RNase E-independent sRNAs. *V. cholerae* wild-type and *rne<sup>TS</sup>* strains were grown at 30°C to early exponential phase ( $OD_{600}$  of 0.2; RyhB and Spot 42) or to stationary phase ( $OD_{600}$  of 2.0; VqmR). Cultures were divided in half and continuously grown at either 30°C or 44°C for 30 min. Cleavage patterns of sRNAs without detectable RNase E cleavage sites were monitored on Northern blots. The genomic locations and relative orientations are shown above the gels. Genes are shown in gray; sRNAs are shown in blue. Filled triangles indicate the size of full-length sRNAs. 5S rRNA served as loading control.



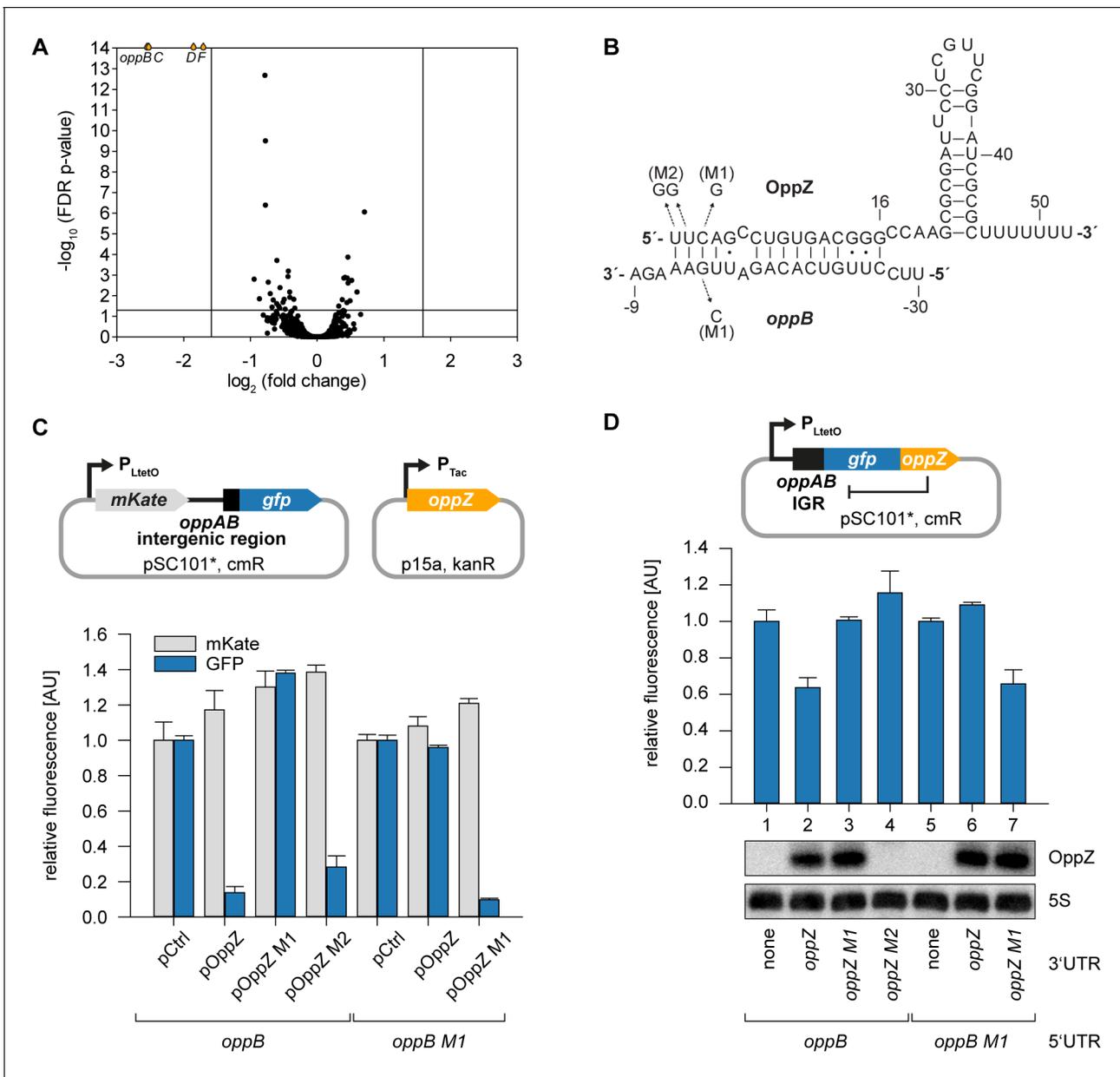
**Figure 2.** OppZ is produced from the *oppABCD* 3' end. (A) Top: Genomic organization of *oppABCD* and *oppZ*. Bottom: Alignment of *oppZ* sequences, including the last codons of *oppF*, from various *Vibrio* species. The *oppF* stop codon, the RNase E cleavage site and the Rho-independent terminator are indicated. (B) *V. cholerae* wild-type and *me<sup>TS</sup>* strains were grown at 30°C to stationary phase (OD<sub>600</sub> of 2.0). Cultures were divided in half and continuously grown at either 30°C or 44°C for 30 min. OppZ synthesis was analyzed by Northern blot with 5S rRNA as loading control. The triangle indicates the size of mature OppZ. (C) Protein and RNA samples were obtained from *V. cholerae oppA::3XFLAG oppB::3XFLAG* strains carrying either the native *oppA* promoter or the inducible pBAD promoter upstream of *oppA*. Samples were collected at the indicated OD<sub>600</sub> and tested for OppA and OppB production by Western blot and for OppZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. Lanes 1–8: Growth without L-arabinose. Lanes 9–12: Growth with either H<sub>2</sub>O (-) or L-arabinose (+) (0.2% final conc.). (D) *V. cholerae* wild-type (control) and *hfq::3XFLAG* (Hfq-FLAG) strains were grown to stationary phase (OD<sub>600</sub> of 2.0), lysed, and subjected to immunoprecipitation using the anti-FLAG antibody. RNA samples of lysate (total RNA) and co-immunoprecipitated fractions were analyzed on Northern blots. 5S rRNA served as loading control.



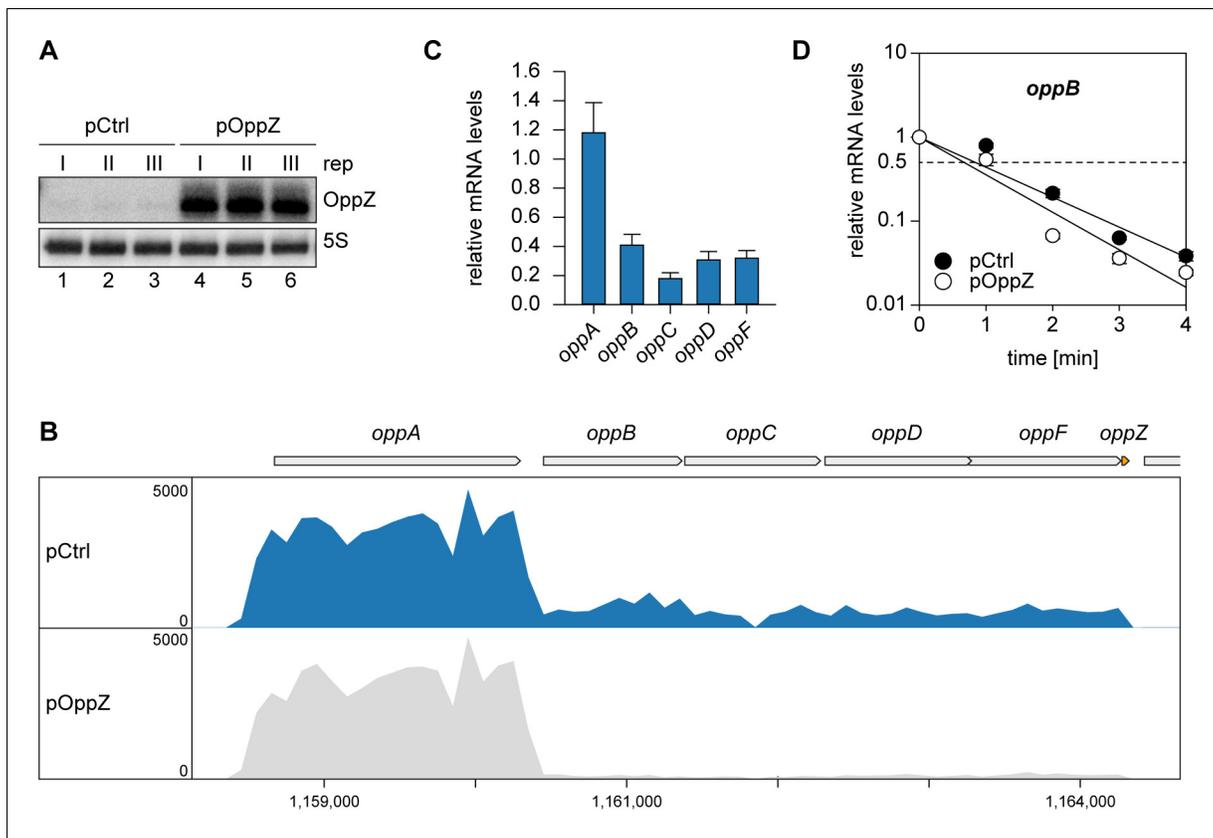
**Figure 2—figure supplement 1.** Hfq dependence of OppZ processing. **(A)** Schematic description of the analyzed OppZ variants. OppZ was produced natively from the genomic *opp* locus, expressed as mature sRNA from a plasmid (pOppZ) or cleaved from a plasmid-encoded precursor transcript including the 3' end of *oppF* (pPrecursor). Expression of both plasmid-based *oppZ* variants was driven by a constitutive promoter. **(B)** *V. cholerae* wild-type,  $\Delta oppZ$ ,  $\Delta hfq$  or  $\Delta hfq \Delta oppZ$  strains carrying *oppA*::3XFLAG *oppB*::3XFLAG genes and a control plasmid or the indicated OppZ expression plasmid were grown to stationary phase (OD<sub>600</sub> of 2.0). RNA samples were collected and OppZ processing was analyzed by Northern blot. 5S rRNA served as loading control.



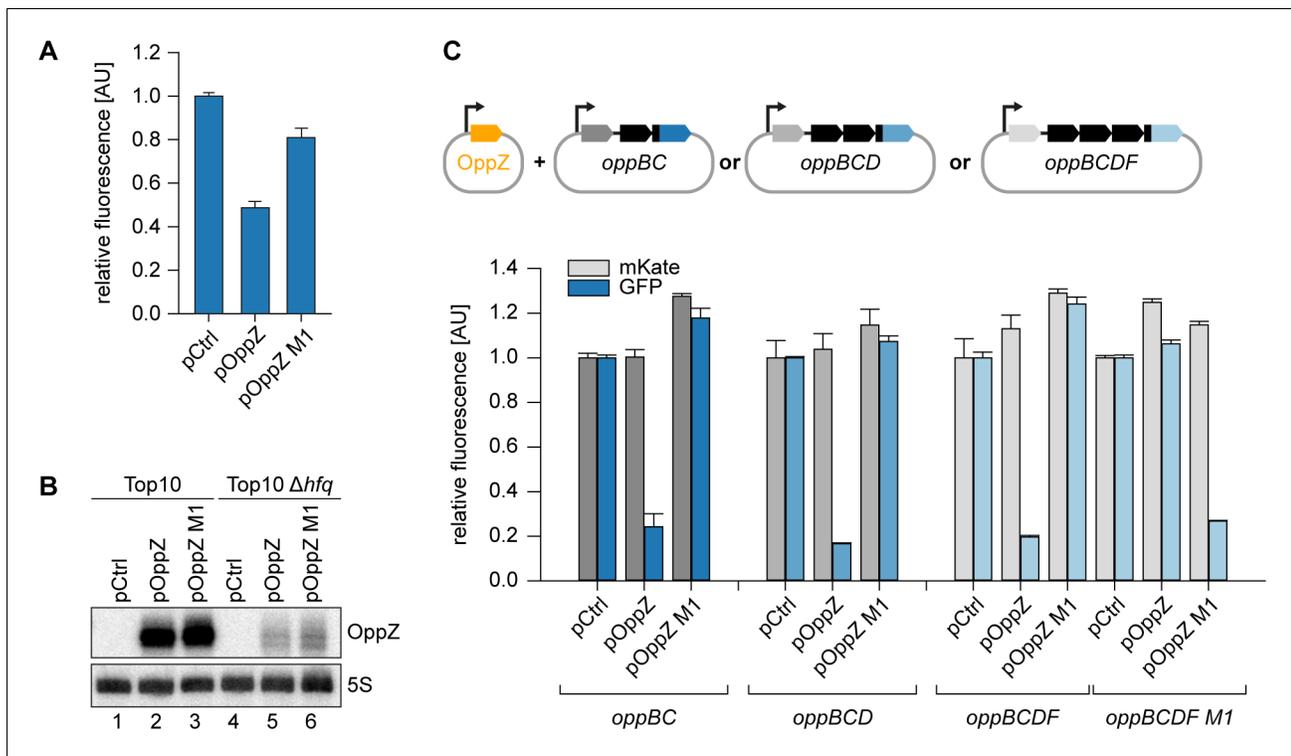
**Figure 2—figure supplement 2.** Hfq dependence of OppZ stability. *V. cholerae* wild-type and  $\Delta hfq$  strains were grown to early stationary phase ( $OD_{600}$  of 1.5) and treated with rifampicin to terminate transcription. RNA samples were obtained at the indicated time points and OppZ transcript levels were monitored by Northern blot and normalized to 5S rRNA levels as loading control. Error bars represent the SD of three biological replicates.



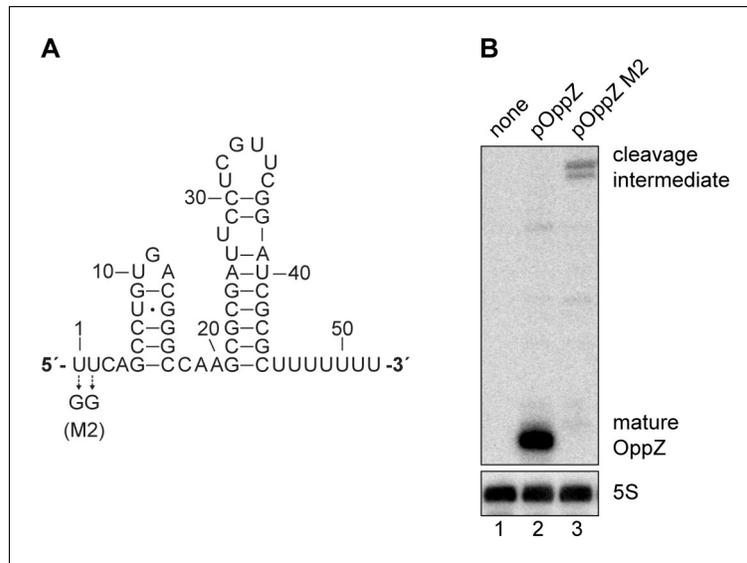
**Figure 3.** Feedback autoregulation at the suboperonic level. (A) Volcano plot of genome-wide transcript changes in response to inducible OppZ over-expression. Lines indicate cut-offs for differentially regulated genes at 3-fold regulation and FDR-adjusted  $p\text{-value} \leq 0.05$ . Genes with an FDR-adjusted  $p\text{-value} < 10^{-14}$  are indicated as droplets at the top border of the graph. (B) Predicted OppZ secondary structure and base-pairing to *oppB*. Arrows indicate the mutations tested in (C) and (D). (C) *E. coli* strains carrying a translational reporter plasmid with the *oppAB* intergenic region placed between *mKate2* and *gfp* were co-transformed with a control plasmid or the indicated OppZ expression plasmids. Transcription of the reporter and *oppZ* were driven by constitutive promoters. Cells were grown to  $\text{OD}_{600} = 1.0$  and fluorophore production was measured. mKate and GFP levels of strains carrying the control plasmid were set to 1. Error bars represent the SD of three biological replicates. (D) Single-plasmid regulation was measured by inserting the indicated *oppZ* variant into the 3' UTR of a translational *oppB::gfp* fusion. Expression was driven from a constitutive promoter. *E. coli* strains carrying the respective plasmids were grown to  $\text{OD}_{600} = 1.0$  and GFP production was measured. Fluorophore levels from control fusions without an sRNA gene were set to one and error bars represent the SD of three biological replicates. OppZ expression was tested by Northern blot; 5S rRNA served as loading control.



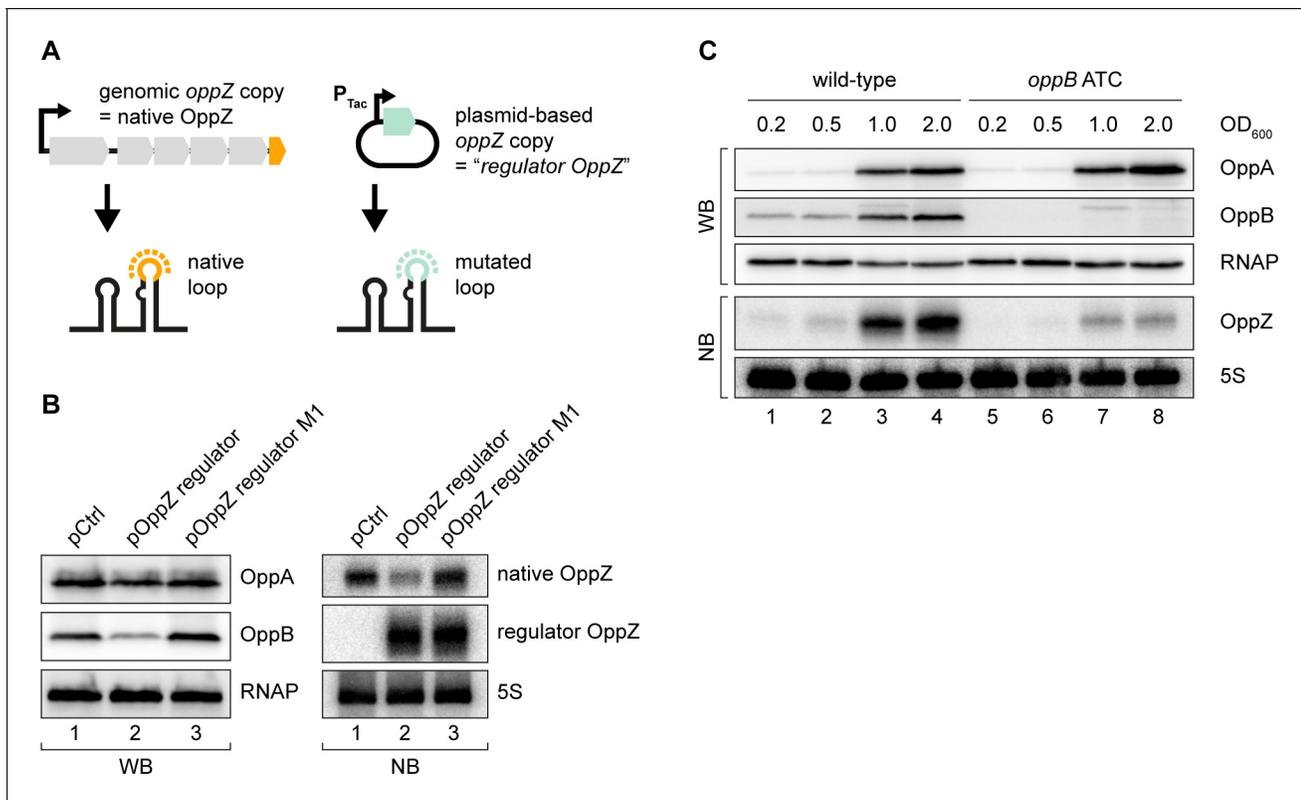
**Figure 3—figure supplement 1.** Pulse expression of OppZ reduces *oppBCDF* transcript levels. (A) *V. cholerae* carrying pBAD1K-*oppZ* (pOppZ) or a control plasmid (pCtrl) were grown in biological triplicates to exponential phase ( $OD_{600}$  of 0.5) and *oppZ* expression was induced by L-arabinose (0.2% final conc.). RNA samples were collected after 15 min and analyzed for OppZ levels by Northern blot; 5S rRNA served as loading control. (B) Samples from (A) were subjected to RNA-seq and average coverage of the *opp* operon is shown for one representative replicate. (C) *V. cholerae*  $\Delta oppZ$  carrying pBAD1K-*oppZ* or a control plasmid were grown to late exponential phase ( $OD_{600}$  of 1.0) and *oppZ* expression was induced by L-arabinose (0.2% final conc.) for 15 min. mRNA levels of *oppABCDF* were analyzed by qRT-PCR. Bars show mRNA levels upon OppZ induction compared to the control; error bars represent the SD of three biological replicates. (D) *V. cholerae*  $\Delta oppZ$  strains carrying either pBAD1K-ctrl (pCtrl) or pBAD1K-*oppZ* (pOppZ) were grown to late exponential phase ( $OD_{600}$  of 1.0) and treated with L-arabinose (0.2% final conc.) to induce sRNA expression. After 15 min of induction, rifampicin was added to terminate transcription. RNA samples were obtained at the indicated time points and *oppB* transcript levels were monitored by qRT-PCR. Error bars represent the SD of three biological replicates.



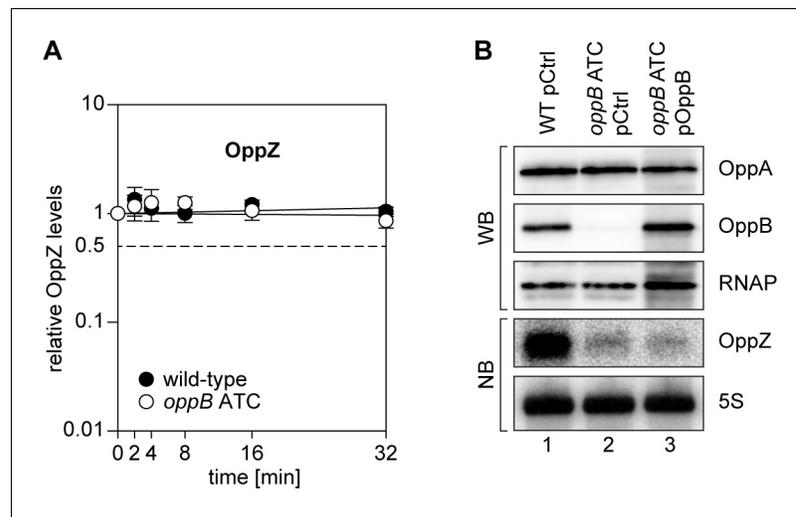
**Figure 3—figure supplement 2.** Hfq-dependent, post-transcriptional repression of OppBCDF by OppZ. (A) *E. coli*  $\Delta hfq$  strains carrying the translational *oppB-gfp* reporter plasmid and either a control plasmid or the indicated OppZ expression plasmids were grown to  $OD_{600} = 1.0$  and fluorophore production was measured. GFP levels of the control strain were set to 1. Error bars represent the SD of three biological replicates. (B) *E. coli* wild-type or  $\Delta hfq$  strains carrying the translational *oppB-gfp* reporter plasmid and either a control plasmid or the indicated OppZ expression plasmids were grown to  $OD_{600} = 1.0$ . RNA samples were analyzed for OppZ levels by Northern blot; 5S rRNA served as loading control. (C) *E. coli* strains carrying translational reporter plasmids with the indicated parts of the *opp* operon placed between *mKate2* and *gfp* were co-transformed with a control plasmid or the respective OppZ expression plasmids. Transcription of the reporter and *oppZ* were driven by constitutive promoters. Cells were grown to  $OD_{600} = 1.0$  and fluorophore production was measured. mKate and GFP levels of strains carrying the control plasmid were set to 1. Error bars represent the SD of three biological replicates.



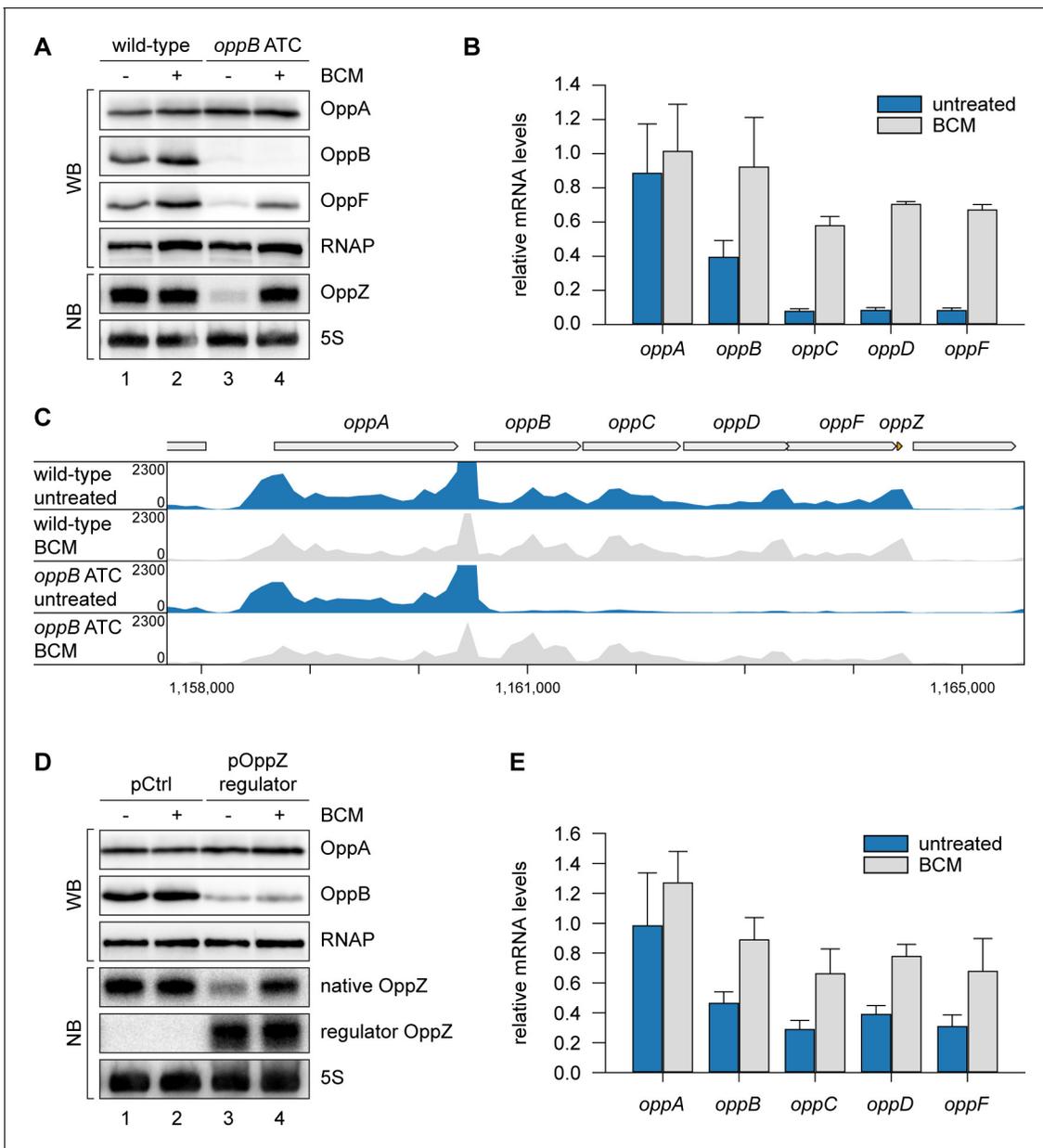
**Figure 3—figure supplement 3.** Mutational analysis of the RNase E site in *oppZ*. (A) Predicted structure of the OppZ sRNA. The M2 mutation blocking cleavage by RNase E is indicated. (B) *E. coli* strains carrying the empty pXG10-SF plasmid or derivatives with the indicated *oppZ* gene in the 3' UTR of *gfp* were grown to  $OD_{600} = 1.0$ . RNA samples were analyzed for OppZ processing by Northern blot; 5S rRNA served as loading control.



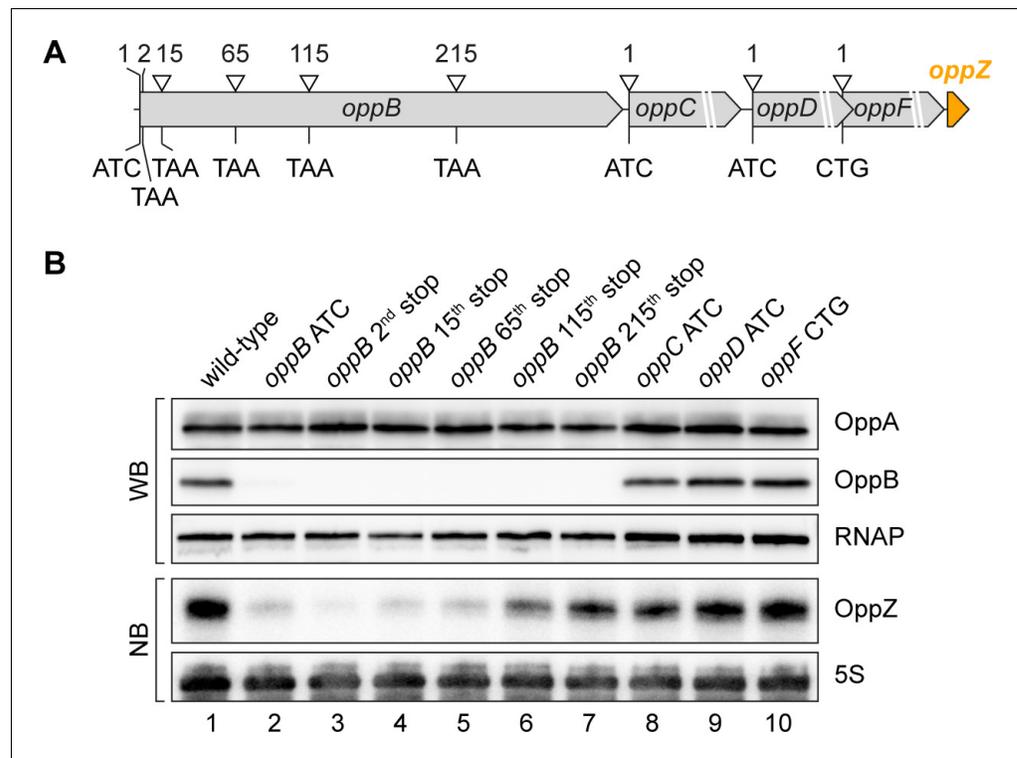
**Figure 4.** Translational control of OppZ synthesis. (A) Schematic of the analyzed OppZ variants containing the native stem loop sequence (produced from the genomic *oppZ* locus) or a mutated stem loop sequence ("regulator OppZ" produced from a plasmid-based constitutive promoter). (B) *V. cholerae oppA::3XFLAG oppB::3XFLAG* carrying a control plasmid (pCMW-1) or a plasmid expressing *regulator OppZ* (pMD194, pMD195) were grown to stationary phase (OD<sub>600</sub> of 2.0). OppA and OppB production were tested by Western blot and expression of native OppZ and regulator OppZ was monitored on Northern blot using oligonucleotides binding to the respective loop sequence variants. RNAP and 5S rRNA served as loading controls for Western blot and Northern blot, respectively. (C) The *oppB* start codon was mutated to ATC in an *oppA::3XFLAG oppB::3XFLAG* background. *V. cholerae* strains with wild-type or mutated *oppB* start codon were grown in LB medium. Protein and RNA samples were collected at the indicated OD<sub>600</sub> and tested for OppA and OppB production by Western blot and for OppZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively.



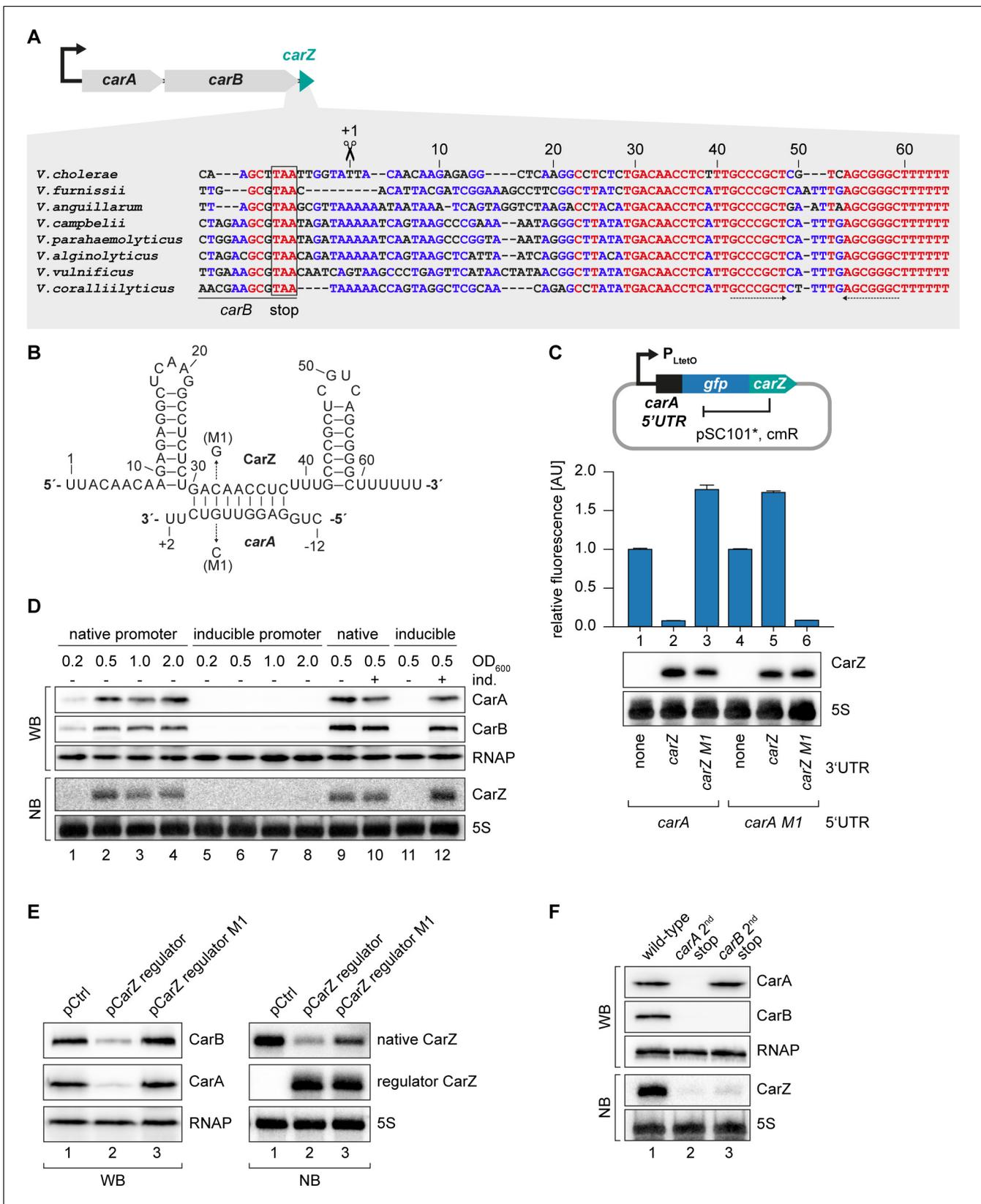
**Figure 4—figure supplement 1.** Translational control of OppZ synthesis. (A) *V. cholerae* wild-type and *oppB* ATC strains were grown to stationary phase ( $OD_{600}$  of 2.0) and treated with rifampicin to terminate transcription. RNA samples were obtained at the indicated time points and OppZ transcript levels were monitored by Northern blot and normalized to 5S rRNA levels as loading control. Error bars represent the SD of three biological replicates. (B) *V. cholerae* wild-type and *oppB* ATC strains carrying either a control plasmid (pCtrl) or an inducible *oppB* complementation plasmid (pOppB) were grown to late exponential phase ( $OD_{600}$  of 1.0) and *oppB* expression was induced by the addition of L-arabinose (0.2% final conc.). Protein and RNA samples were obtained after 60 min and tested for OppA and OppB production by Western blot and for OppZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively.



**Figure 5.** OppZ promotes transcription termination through Rho. (A) *V. cholerae* *oppA*::3XFLAG *oppB*::3XFLAG *oppF*::3XFLAG strains with wild-type or mutated *oppB* start codon were grown to early stationary phase (OD<sub>600</sub> of 1.5). Cultures were divided in half and treated with either H<sub>2</sub>O or BCM (25 μg/ml final conc.) for 2 hr before protein and RNA samples were collected. OppA, OppB and OppF production were tested by Western blot and OppZ expression was monitored by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. (B) Biological triplicates of *V. cholerae* *oppA*::3XFLAG *oppB*::3XFLAG strains with wild-type or mutated *oppB* start codon were treated with BCM as described in (A). *oppABCDF* expression in the *oppB* start codon mutant compared to the wild-type control was analyzed by qRT-PCR. Error bars represent the SD of three biological replicates. (C) Triplicate samples from (B) were subjected to Term-seq and average coverage of the *opp* operon is shown for one representative replicate. The coverage cut-off was set at the maximum coverage of annotated genes. (D) *V. cholerae* *oppA*::3XFLAG *oppB*::3XFLAG strains carrying a control plasmid (pMD397) or a plasmid expressing regulator *OppZ* (pMD398) were treated with BCM as described in (A). OppA and OppB production were tested by Western blot and expression of native OppZ and regulator OppZ was monitored on Northern blot using oligonucleotides binding to the respective loop sequence variants. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. (E) Levels of *oppABCDF* in the experiment described in (D) were analyzed by qRT-PCR. Error bars represent the SD of three biological replicates.



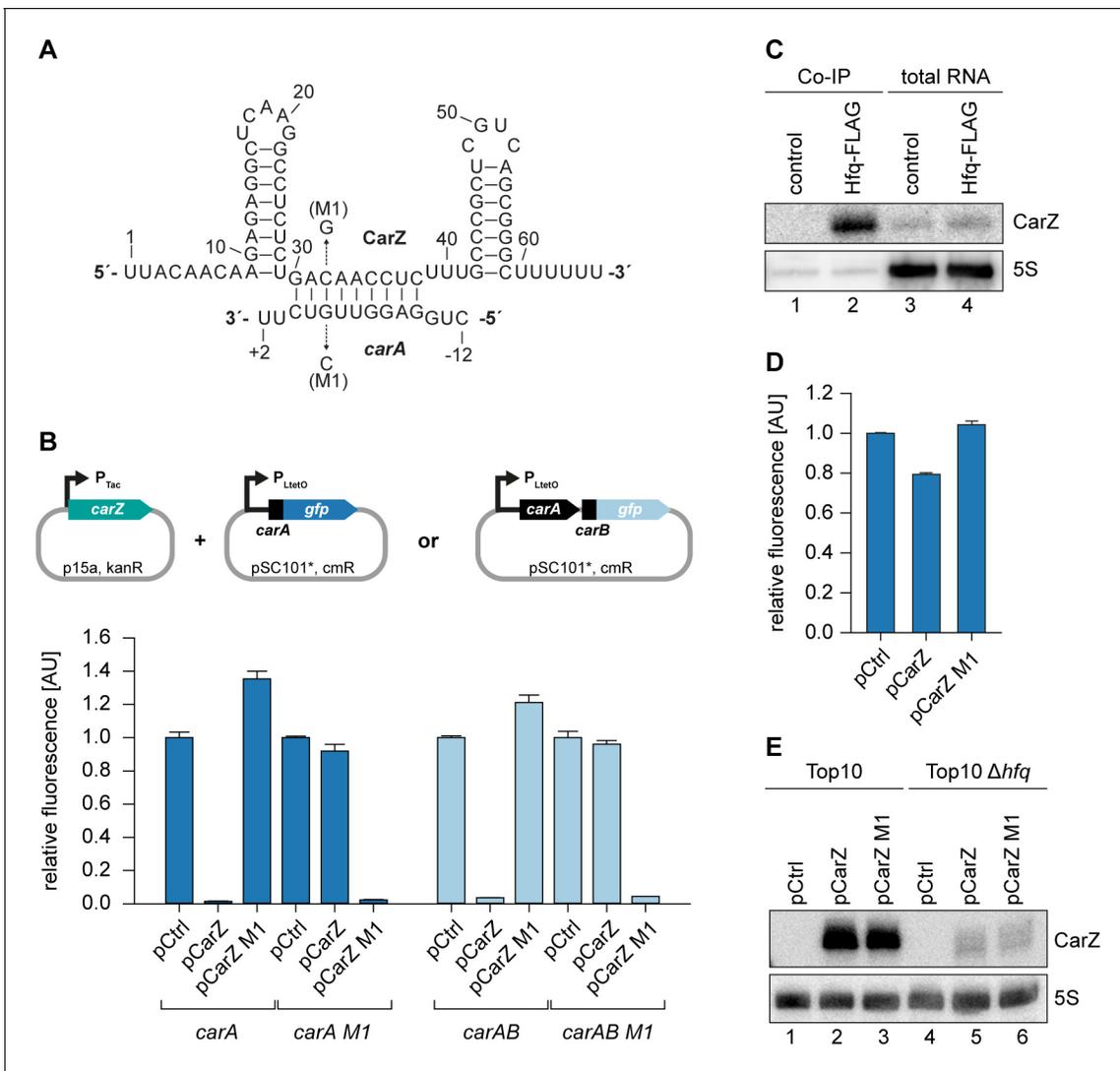
**Figure 6.** Influence of OppBCDF translation on OppZ expression. (A) The depicted mutations were individually inserted into the *opp* locus to inactivate the start codons of *oppB*, *oppC*, *oppD* or *oppF* or to insert STOP codons at the positions 2, 15, 65, 115 or 215 of *oppB*. (B) *V. cholerae oppA::3XFLAG oppB::3XFLAG* strains with the described *opp* mutations were grown: wild-type (lane 1), the *oppB* start codon mutated (lane 2), a STOP codon inserted at the 2<sup>nd</sup>, 15<sup>th</sup>, 65<sup>th</sup>, 115<sup>th</sup> or 215<sup>th</sup> codon of *oppB* (lanes 3–7) or mutated start codons of *oppC*, *oppD* or *oppF* (lanes 8–10). At stationary phase (OD<sub>600</sub> of 2.0), protein and RNA samples were collected and tested for OppA and OppB production by Western blot and for OppZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively.



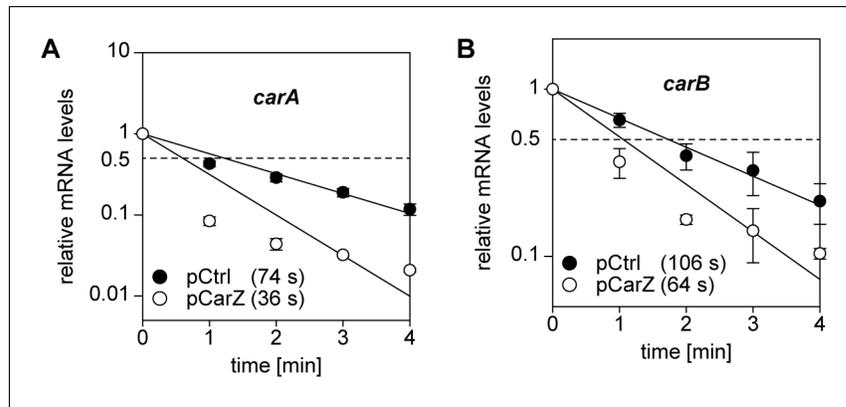
**Figure 7.** CarZ is another autoregulatory sRNA from *V. cholerae*. (A) Top: Genomic context of *carAB* and *carZ*. Bottom: Alignment of *carZ* sequences, including the last codons of *carB*, from various *Vibrio* species. The *carB* stop codon, the RNase E cleavage site and the Rho-independent terminator are indicated. Figure 7 continued on next page

## Figure 7 continued

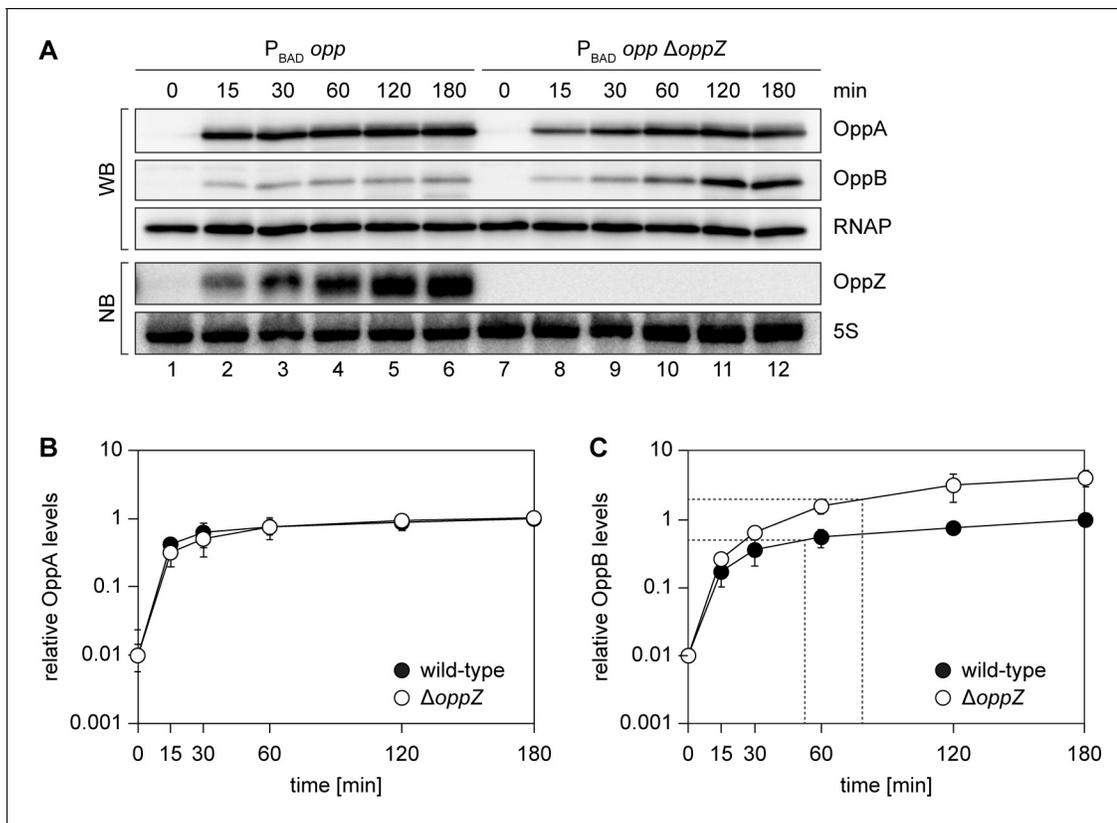
indicated. (B) Predicted CarZ secondary structure and base-pairing to *carA*. Arrows indicate the single nucleotide mutations tested in (C). (C) Single-plasmid feedback regulation of *carA* by CarZ was measured by inserting the indicated *carZ* variant into the 3' UTR of a translational *carA::gfp* fusion. Expression was driven from a constitutive promoter. *E. coli* strains carrying the respective plasmids were grown to  $OD_{600} = 1.0$  and GFP production was measured. Fluorophore levels from control fusions without an sRNA gene were set to one and error bars represent the SD of three biological replicates. CarZ expression was tested by Northern blot; 5S rRNA served as loading control. (D) Protein and RNA samples were obtained from *V. cholerae carA::3XFLAG carB::3XFLAG* carrying either the native *carA* promoter or the inducible pBAD promoter upstream of *carA*. Samples were collected at the indicated  $OD_{600}$  and tested for CarA and CarB production by Western blot and for CarZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. Lanes 1–8: Growth without L-arabinose. Lanes 9–12: Growth with either H<sub>2</sub>O (-) or L-arabinose (+) (0.2% final conc.). (E) *V. cholerae carA::3XFLAG carB::3XFLAG* strains carrying a control plasmid or a plasmid expressing a CarZ variant with a mutated stem loop (*regulator CarZ*) were grown to late exponential phase ( $OD_{600}$  of 1.0). CarA and CarB production were tested by Western blot and expression of native CarZ or *regulator CarZ* was monitored on Northern blot using oligonucleotides binding to the respective loop sequence variants. RNAP and 5S rRNA served as loading controls for Western blot and Northern blot, respectively. (F) *V. cholerae carA::3XFLAG carB::3XFLAG* strains with the following *carA* or *carB* mutations were grown: wild-type (lane 1) or a STOP codon inserted at the 2<sup>nd</sup> codon of *carA* (lane 2) or *carB* (lane 3), respectively. At late exponential phase ( $OD_{600}$  of 1.0), protein and RNA samples were collected and tested for CarA and CarB production by Western blot and for CarZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively.



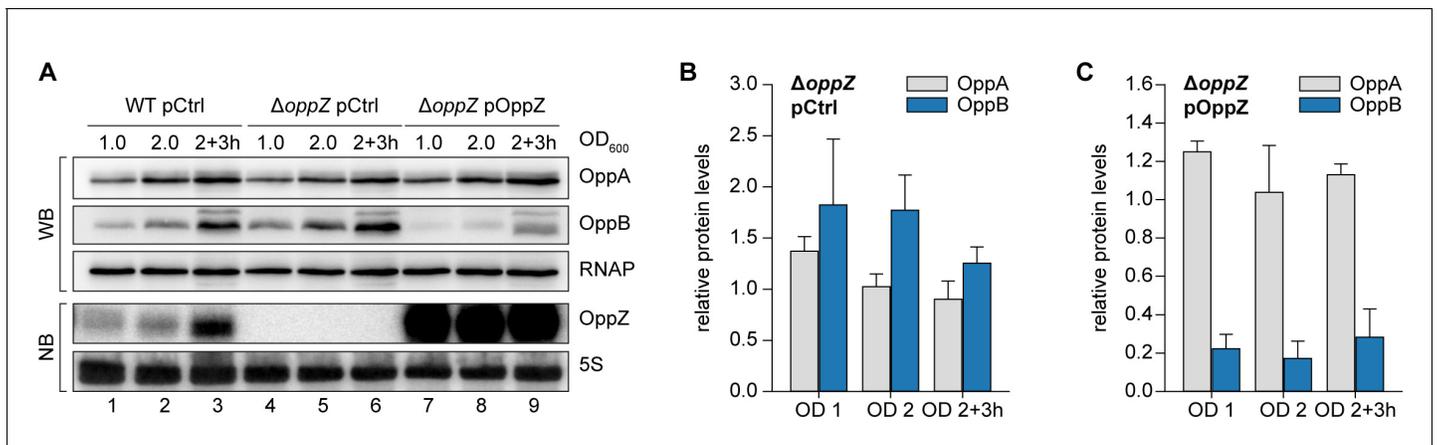
**Figure 7—figure supplement 1.** Hfq-dependent, post-transcriptional repression of CarA and CarB by CarZ. (A) Predicted CarZ secondary structure and base-pairing to *carA*. Arrows indicate the single nucleotide mutations tested in (B). (B) *E. coli* strains carrying translational reporter plasmids for *carA::gfp* or *carAB::gfp* were co-transformed with a control plasmid or the indicated CarZ expression plasmids. Transcription of the reporter and *carZ* were driven by constitutive promoters. Cells were grown to  $OD_{600} = 1.0$  and fluorophore production was measured. GFP levels of strains carrying the control plasmid were set to 1. Error bars represent the SD of three biological replicates. (C) *V. cholerae* wild-type (control) and *hfq::3XFLAG* (Hfq-FLAG) strains were grown to stationary phase ( $OD_{600}$  of 2.0), lysed, and subjected to immunoprecipitation using the anti-FLAG antibody. RNA samples of lysate (total RNA) and co-immunoprecipitated fractions were analyzed on Northern blots. 5S rRNA served as loading control. (D) *E. coli*  $\Delta hfq$  strains carrying the translational *carA::gfp* reporter plasmid and either a control plasmid or the indicated CarZ expression plasmids were grown to  $OD_{600} = 1.0$  and fluorophore production was measured. GFP levels of the control strain were set to 1. Error bars represent the SD of three biological replicates. (E) *E. coli* wild-type or  $\Delta hfq$  strains carrying the translational *carA::gfp* reporter plasmid and either a control plasmid or the indicated CarZ expression plasmids were grown to  $OD_{600} = 1.0$ . RNA samples were analyzed for CarZ levels by Northern blot; 5S rRNA served as loading control.



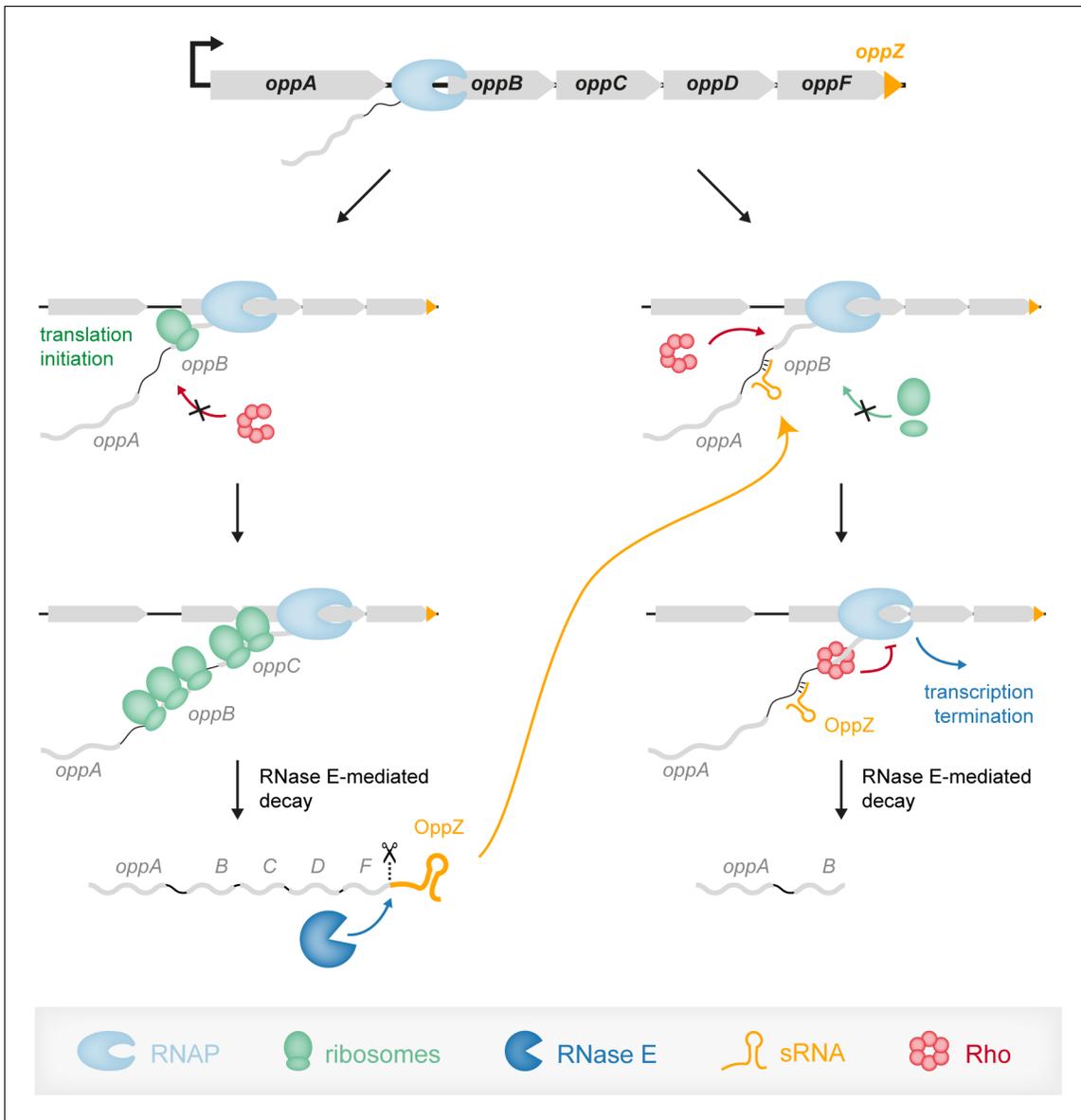
**Figure 7—figure supplement 2.** CarZ induces *carAB* degradation. (A, B) *V. cholerae*  $\Delta carZ$  strains carrying either pBAD1K-ctrl (pCtrl) or pBAD1K-*carZ* (pCarZ) were grown to late exponential phase ( $OD_{600}$  of 1.0) and treated with L-arabinose (0.2% final conc.) to induce sRNA expression. After 15 min of induction, rifampicin was added to terminate transcription. RNA samples were obtained at the indicated time points and transcript levels of *carA* (A) and *carB* (B) were monitored by qRT-PCR. Error bars represent the SD of three biological replicates.



**Figure 8.** Modified kinetics of gene induction by autoregulatory OppZ. (A) Expression of the *opp* operon including the *oppA*::3XFLAG and *oppB*::3XFLAG genes and the native *oppZ* gene (lanes 1–6) or an *oppZ* deletion (lanes 7–12) was induced from the pBAD promoter at late exponential phase ( $OD_{600}$  of 1.0) by the addition of L-arabinose (0.2% final conc.). Protein and RNA samples were obtained at the indicated time points and tested for OppA and OppB production by Western blot and for OppZ expression by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. (B, C) Quantification of OppA (B) or OppB (C) levels from the experiment in (A); error bars represent the SD of three biological replicates. Data are presented as fold regulation of OppA or OppB in  $\Delta oppZ$  compared to the wild-type. Dashed lines in (C) indicate the time points of half-maximum OppB expression.



**Figure 8—figure supplement 1.** OppZ-dependent repression of OppA and OppB protein levels. (A) *V. cholerae* wild-type and  $\Delta oppZ$  strains carrying the *oppA*::3XFLAG and *oppB*::3XFLAG genes and either a control plasmid or a constitutive OppZ expression plasmid were grown to obtain protein and RNA samples at the indicated OD<sub>600</sub>. OppA and OppB production were analyzed by Western blot and OppZ expression was tested by Northern blot. RNAP and 5S rRNA served as loading controls for Western and Northern blots, respectively. (B) Quantification of (A), bars show fold regulation of OppA and OppB in  $\Delta oppZ$  compared to the wild-type; error bars represent the SD of four biological replicates. (C) Quantification of (A), bars show fold regulation of OppA and OppB upon OppZ overexpression in the  $\Delta oppZ$  background compared to the wild-type control; error bars represent the SD of four biological replicates.



**Figure 9.** Model of the OppZ-dependent mechanism of *opp* regulation. Transcription of the *oppABCDF* operon initiates upstream of *oppA* and in the absence of OppZ (left) involves all genes of the operon as well as OppZ. In this scenario, all cistrons of the operon are translated. In the presence of OppZ (right), the sRNA blocks translation of *oppB* and the ribosome-free mRNA is recognized by termination factor Rho. Rho catches up with the transcribing RNAP and terminates transcription pre-maturely within *oppB*. Consequently, *oppBCDF* are not translated and OppZ is not produced.

## 4 Symbiosis, virulence and natural products biosynthesis in entomopathogenic bacteria are regulated by a small RNA

Nick Neubacher\*, Nicholas J. Tobias\*, **Michaela Huber\***, Xiaofeng Cai, Timo Glatter, Sacha J. Pidot, Timothy P. Stinear, Anna Lena Lütticke, Kai Papenfort, and Helge B. Bode (2020). “Symbiosis, virulence and natural products biosynthesis in entomopathogenic bacteria are regulated by a small RNA.”

\* These authors contributed equally.

Manuscript accepted for publication in *Nature Microbiology*.

(*Letter* format)

Published version:

Neubacher N, Tobias NJ, Huber M *et al.* (2020) Symbiosis, virulence and natural-product biosynthesis in entomopathogenic bacteria are regulated by a small RNA. *Nat Microbiol* 5, 1481–1489.

[doi.org/10.1038/s41564-020-00797-5](https://doi.org/10.1038/s41564-020-00797-5)

<https://www.nature.com/articles/s41564-020-00797-5>



## **Symbiosis, virulence and natural products biosynthesis in entomopathogenic bacteria are regulated by a small RNA**

Nick Neubacher<sup>1\*</sup>, Nicholas J. Tobias<sup>1,2,3\*</sup>, Michaela Huber<sup>4,5,6\*</sup>, Xiaofeng Cai<sup>1</sup>, Timo Glatter<sup>7</sup>, Sacha J. Pidot<sup>8</sup>, Timothy P. Stinear<sup>8</sup>, Anna Lena Lütticke<sup>1</sup>, Kai Papenfort<sup>4,5</sup> and Helge B. Bode<sup>1,2,3,9,#</sup>

<sup>1</sup> Molekulare Biotechnologie, Fachbereich Biowissenschaften, Goethe-Universität Frankfurt, Frankfurt am Main, Germany

<sup>2</sup> LOEWE Center for Translational Biodiversity in Genomics (TBG), Germany

<sup>3</sup> Senckenberg Gesellschaft für Naturforschung, Frankfurt am Main, Germany

<sup>4</sup> Friedrich Schiller University Jena, Institute of Microbiology, 07745 Jena, Germany

<sup>5</sup> Microverse Cluster, Friedrich Schiller University Jena, 07743 Jena, Germany

<sup>6</sup> Ludwig-Maximilians-University of Munich, Faculty of Biology I, 82152 Martinsried, Germany

<sup>7</sup> Core Facility for Mass Spectrometry and Proteomics, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

<sup>8</sup> Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia

<sup>9</sup> Buchmann Institute for Molecular Life Sciences, Goethe-Universität Frankfurt, Frankfurt am Main, Germany

\*Co-first authors

#Corresponding author

[h.bode@bio.uni-frankfurt.de](mailto:h.bode@bio.uni-frankfurt.de)

*Photorhabdus* and *Xenorhabdus* species have mutualistic associations with nematodes and an entomopathogenic stage (1, 2) in their lifecycles. In both stages numerous specialized metabolites are produced that have roles in symbiosis and virulence (3, 4). Although regulators have been implicated in the regulation of these specialized metabolites (3, 4), how small regulatory RNAs (sRNAs) are involved in this process is not clear. We show here that the Hfq-dependent sRNA, ArcZ, is required for specialized metabolite production in *Photorhabdus* and *Xenorhabdus*. We discovered that ArcZ directly base-pairs with the mRNA encoding HexA, which represses the expression of specialized metabolite gene clusters. In addition to specialized metabolite genes, we show that the ArcZ regulon affects ~15% of all transcripts in *Photorhabdus* and *Xenorhabdus*. Thus, the ArcZ sRNA is crucial for specialized metabolite production in *Photorhabdus* and *Xenorhabdus* species and might well become a useful tool for metabolic engineering and identification of commercially relevant natural products.

Regulation via *trans*-encoded sRNAs typically occurs by imperfect base-pairing of sRNAs with their mRNA targets and can be mediated by RNA chaperones such as Hfq and ProQ (5, 6). RNA duplex formation is usually short (6 to 12 nucleotides) and can result in conformational changes in RNA secondary structure with various regulatory outcomes (7). The RNA chaperone Hfq is highly conserved throughout the bacterial kingdom (8). Several complex phenotypes have been attributed to Hfq with its regulatory roles being achieved by stabilizing sRNAs and/or mRNAs, mediating base-pairing of sRNAs and their targets, modulation of mRNA translation (8), as well as accelerating the degradation of sRNAs and their targets (5). Expression of sRNAs is highly dynamic, with sRNA profiles in *Salmonella* shown to be strongly dependent on the bacterial growth phase (9). ArcZ is one of the few Hfq-bound sRNAs whose expression remains relatively constant in *Salmonella* throughout the growth phases, making up ~7-12% of all reads identified by Hfq co-immunoprecipitation experiments (9). ArcZ is transcribed as a 129 nt primary transcript (Figure 1a) and processed into a stable short form (~50 nt) (9-11). The processed short form of ArcZ directly activates *rpoS* translation and inhibits the expression of several other genes (11, 12). In *E. coli*, the expression of *arcZ* is repressed by the ArcA-ArcB two-component system under anaerobic conditions. In a negative feedback loop, *arcZ* represses, and is repressed by *arcB* transcription (11). Although there is a wealth of research on ArcZ in *E. coli* and *Salmonella* (9-11), its function in other bacteria remains unclear.

SM in bacteria are often responsible for ecologically important activities (13). In the case of *Xenorhabdus* and *Photorhabdus*, SMs play an essential role in cross-kingdom interactions with nematodes, various insects, as well as bacterial and fungal species competing for the same food source (14). Our earlier work on *Photorhabdus* showed that deletion of *hfq* resulted in severe perturbation of gene networks, including several key regulators (3). This led to an overall decrease in SM production and a failure of the bacteria to support their obligate symbiosis with nematodes. Despite SMs playing a central role in the life cycle of the symbiosis, the exact ecological function for many of these compounds remained unknown. Over the past years, significant advances have been made towards finding bioactivities for many of the SMs, with assigned functions including cell-cell communication (photopyrones, dialkylresorcinols (15, 16)), nematode development (isopropylstilbene (17)), defense against food-competitors (isopropylstilbene, rhabdopeptides (17, 18)), or insect pathogenicity (rhabduscin, rhabdopeptides, glidobactin (18-20)). However, understanding the full potential of SMs in these bacteria is still hampered by a somewhat limited understanding of when individual SMs are produced, and their regulation in general. Regulation of SM in *Photorhabdus* and *Xenorhabdus* so far implicated the regulators Hfq, HexA (also LrhA), LeuO and Lrp (3, 4, 21, 22). Deletion of *hfq* in *Photorhabdus* resulted in complex regulatory changes, including a strong up-regulation of

HexA, a known repressor of SM production (4). Consequently, SM production was completely abolished in this strain and nematode development was severely restricted.

Given the overlapping lifecycles and niche occupation, we hypothesized that deletion of *hfq* in *Xenorhabdus* would have a similar effect on the production of SMs and the transcriptome. We confirmed this in *X. szentirmaii* DSM16338 using both high-performance LCMS/MS and RNA-seq (Supplementary Note 1, Supplementary Table 1). To further elucidate the mechanism of SM regulation, we investigated Hfq binding partners. To this end, we sequenced both *X. szentirmaii* and *P. laumondii* using a sRNA sequencing protocol and combined this with CappableSeq data to globally annotate transcriptional start sites belonging to coding sequences or potential previously undescribed sRNAs (Supplementary Note 2, Supplementary Table 2). We confirmed expression of several of these sRNAs by Northern blot analysis (Extended Data Figures 1 & 2). To identify RNA-protein interactions on a global scale, we next employed RNA immunoprecipitation followed by high-throughput sequencing (RIP-seq) using chromosomally produced Hfq::3xFLAG protein as bait. We performed these experiments at two different cell densities (*i.e.* OD<sub>600</sub> 0.5 and OD<sub>600</sub> 5.0, for a full list of ODs from different experiments, see Supplementary Table 3). From the corresponding sequencing data, we first identified regions of 5 bp or more that were enriched in our tagged Hfq strain (see Methods). We then searched for sRNAs that were specifically enriched in the tagged samples, when compared to the untagged samples. We identified a total of 37 binding sites in annotated sRNAs (35 unique sRNAs) at OD<sub>600</sub> 0.5 and 37 binding sites (34 unique) at OD<sub>600</sub> 5.0 that were enriched by at least three-fold in both replicates (Figure 1b, Supplementary Table 4). During early exponential growth, 11 sRNAs (out of 35) were identified that are described to associate with Hfq in other species, while 10 (out of 34) are known from those that were enriched at OD<sub>600</sub> 5.0. As a second step, we examined mRNAs enriched in the data. At OD 5.0, 402 mRNAs and 32 annotated 5' UTRs were identified to associate with Hfq. At OD 0.5 a total of 1,003 mRNAs and 29 5' UTRs were detected (Figure 1c, Supplementary Table 5).

We hypothesized that the performed Hfq RIP-seq analysis would allow us to identify key sRNAs involved in SM repression. However, our analysis identified >50 potential sRNAs binding Hfq (across both ODs, Figure 1b). Therefore, rather than individually deleting each sRNA, we constructed a transposon mutant library using pSAM-BT\_Kan (see Methods and Supplementary Note 3) and searched for phenotypes consistent with that of the  $\Delta hfq$  strain. The red color afforded to the bacteria by anthraquinone (AQ) production makes the strain especially suitable for transposon mutagenesis when screening for mutants defective in SM biosynthesis. We screened approximately 60,000 clones for obvious phenotypic alterations. Several mutants were defective in some facets of SM production and showed growth defects (Supplementary Note 3, Extended Data Figure 3, Supplementary Table 6), however, only

one displayed the desired phenotype. Re-sequencing of this strain followed by read mapping revealed that the transposon was inserted within an intergenic region associated with the *arcZ* sRNA gene (Supplementary Figure 1).

ArcZ is a well-known Hfq-associated sRNA, which also appeared in our list of Hfq-bound sRNAs in *P. laumondii* (Figure 1b, Supplementary Table 4). To verify that the observed phenotype was derived from the transposon-insertion, we generated a  $\Delta arcZ$  mutant by deleting the major part of the sRNA (Supplementary Figure 1) and a complemented strain by reintroducing an intact version of *arcZ* at the original locus. Northern blot analysis was performed to verify the absence of ArcZ in the deletion mutant and the presence of ArcZ in the WT and the complementation mutant (Figure 2a). RNA sequencing of  $\Delta arcZ$  mutant showed severe transcriptomic changes compared to the WT and  $\Delta arcZ::arcZ$  mutant of *P. laumondii*, reminiscent of that seen in *P. laumondii*  $\Delta hfq$  (Supplementary Note 4, Supplementary Tables 7 & 8, Supplementary Figure 2). SM production titers in the  $\Delta arcZ$  strain were strongly decreased, similar to that seen in the transposon-insertion mutant and the complementation strain restored SM production (Figure 2b-h).

To corroborate the role of ArcZ in SM production, ArcZ mRNA targets were predicted using CopraRNA (23) (Supplementary Tables 9 & 10). One hit, warranting further investigation was *hexA* (*lrhA*), which was previously identified as a highly upregulated gene in our strains and which represses SM production in both *P. laumondii* (24) and *Xenorhabdus* (3). CopraRNA predicted a 9 bp-long RNA duplex involving the 5' UTR of *hexA* and the processed isoform of ArcZ (Figure 3a). This base-pairing is reminiscent of previously reported ArcZ targets in other bacteria requiring RNase E-mediated release of the sRNA's seed region (12). We also identified a corresponding enriched RNA sequence upstream of the *hexA* CDS at OD<sub>600</sub> 0.5 in the RIP-seq experiments (Figure 1c, Supplementary Figure 3). We hypothesized that, through Hfq, ArcZ might bind to the *hexA* transcript leading to repression of HexA. In lab cultures, where SMs are produced, we hypothesized that Hfq and ArcZ prevent HexA production, allowing the strain to synthesize SM. However, if either *hfq* or *arcZ* were deleted, we would expect that *hexA* is no longer repressed, resulting in severely reduced SM production. To test this idea, we altered the predicted site of the ArcZ-*hexA* interaction to a *PacI* restriction site (TTAATTAA) and created a knock-in of *hexA* with the modified sequence in a  $\Delta hexA$  strain (Supplementary Figure 4a & b). We predicted that a knock-in of *hexA* with an altered 5' UTR would result in a failure of ArcZ to bind, leading to reduced SM titers. Indeed, the SM production titers in the knock-in mutant with the altered binding site upstream of *hexA* were greatly reduced (Figure 2c-h).

To verify the proposed interaction region, we conducted a compensatory base mutation study in *E. coli*. The fifth base-pair of the proposed interaction region was exchanged in the *arcZ* sequence, the *hexA* 5' UTR, or both by site directed mutagenesis (Figure 3a). The *hexA* 5' UTR sequence was fused to *gfp*. The GFP output was measured to determine the efficiency of inhibition (Figure 3b & c). For the control, the GFP signal derived from the expression of *hexA::gfp* was measured and set to 1. When p-*arcZ* was expressed together with *hexA::gfp*, HexA repression was increased 32-fold compared to the control. Additionally, when p-*arcZ*\* (G79C) was expressed, ArcZ\* was no longer able to repress HexA. For *hexA\*::gfp* (C-46G) in combination with the native ArcZ, HexA repression was only slightly increased compared to the control, suggesting that ArcZ can still bind to the 5' UTR of *hexA* but with a much reduced efficiency. When combining p-*arcZ*\* (G79C) with *hexA\*::gfp* (C-46G), HexA::GFP repression was increased 39-fold, which confirms our hypothesis that ArcZ binds to the 5'-UTR of *hexA* to repress HexA production. Of note, this base-pairing sequence is located ~50 nts upstream of the *hexA* translational start site (Fig. 3a) and thus ArcZ binding is unlikely to compete with recognition of the mRNA by 30S ribosomes (25). Instead, alignment of the *P. laumondii hexA* 5' UTR revealed that the ArcZ binding site is CA-rich and highly conserved among other SM-producing bacteria (Supplementary Figure 5). CA-rich sequences located in proximity to translation initiation sites are well-known translational enhancers and sequestration of these regulatory elements by sRNAs has been reported to down-regulate gene expression (26, 27), which might also be relevant for the ArcZ-*hexA* interaction reported here. In addition, we conducted a proteomic analysis with the WT,  $\Delta arcZ$ ,  $\Delta hfq$  and  $\Delta hexA::hexA\_PacI\_UTR$  strains of *P. laumondii*. We used a label free quantification of quadruplicate samples to determine the HexA abundance in each strain. HexA levels were significantly elevated in all mutant strains (11.8 to 22.7 fold, Supplementary Table 11) compared to the WT, further supporting this mechanism of regulation for SM production.

The *arcZ* gene and its genomic organization are highly conserved among enterobacterial species (10) (Figure 1a). Since the control of SMs in *Photothabdus* relays a fundamental ability for these bacteria to occupy their specific niche, we investigated the possibility that the same mechanism occurs in the closely related *Xenorhabdus*. Given the SM reduction in *X. szentirmaii*  $\Delta hfq$ , we constructed a  $\Delta arcZ$  mutant in *X. szentirmaii* in a similar fashion to *P. laumondii*, by deleting 90bp of the predicted *arcZ* sequence. We verified via Northern blots that ArcZ was no longer produced by the deletion mutant and that complementation of the deletion led to production of ArcZ again (Figure 3d). Subsequently, we investigated the transcriptome and SM profile of the WT, deletion and complementation mutant. (Figure 3e, Supplementary Table 12). Consistent with *P. laumondii*, deletion of *arcZ* resulted in a global effect on the transcriptome as well as severely reduced SM titers, both

of which was complemented in the  $\Delta arcZ::arcZ$  complementation mutant (Figure 3e, Figure 4).

Our results highlight the critical role of ArcZ in regulating specialized metabolism in these strains. In fact, the critical nature of SM from *Photorhabdus* and *Xenorhabdus* in modulating the insect immune response indicated that ArcZ might be required for niche occupation by these bacteria. In the *P. laumondii*  $\Delta arcZ$  strain, we observed an inability to support nematode development (Extended Data Figure 4), consistent with our earlier observations in the  $\Delta hfq$  mutant (4). However, the same was not seen in *X. szentirmaii*. We suspect this might be because of the observed increase in protoporphyrin IX (PPIX) production in the *X. szentirmaii*  $\Delta arcZ$  strain (Supplementary Note 5). PPIX is a precursor of heme, which is an important cofactor for key biological processes such as oxidative metabolism (28), protein translation (29), maintaining protein stability (30) and many others. However, PPIX cannot be synthesized *de novo* by *Caenorhabditis elegans* and other nematodes (31). The nematodes therefore rely on external PPIX sources (such as from symbiotic bacteria), which positively affects their growth, reproduction and development (32). It is interesting that despite *P. laumondii* also being capable of producing PPIX, the *Heterorhabditis* nematode reproduction was not supported in either the  $\Delta arcZ$  mutant, nor the  $\Delta hfq$  strain. This is possibly indicative of the nematode specific requirements for reproduction, which may also include isopropylstilbene as an essential factor in *Heterorhabditis* (17), where no analogous compound is yet known to be required for *Steinernema*.

In both *Xenorhabdus* and *Photorhabdus*, nearly all analyzed SM-related genes were found to be down-regulated in the  $\Delta arcZ$  mutant, in accordance with the impaired SM production (Figure 4a & b). This provides a chemical background that is devoid of natural products, which allows for isolation and identification of a desired compound due to the absence of compounds with similar retention times. Therefore,  $\Delta arcZ$  mutants could offer a powerful tool for (over-) production and identification of previously undescribed natural products. As a proof of concept, we conducted a promoter exchange in front of *gxpS* in both *X. szentirmaii*  $\Delta arcZ$  and *X. szentirmaii*  $\Delta hfq$  and compared GXP-C production after induction to the WT (Figure 4c). GXP-C production was found to be increased 90.4 ( $\pm 4.7$ )-fold in *X. szentirmaii*  $\Delta arcZ::pCEP\_GxpS$  and increased 138.6 ( $\pm 17.1$ )-fold in *X. szentirmaii*  $\Delta hfq::pCEP\_GxpS$  compared to the WT (Figure 4c). The striking increase in production, as well as the dramatically reduced chemical background in both strains, highlights the potential of exploiting this regulatory cascade for selective SM production in a strain well-suited for natural product detection. Recently, we showed that this strategy could be applied in a high-throughput manner for rapid screening of bioactivities (33). The same strategy used here in a  $\Delta arcZ$  strain, demonstrates an alternative route to activation, without the complex

perturbations associated with deleting the major RNA chaperone in these bacteria. Interestingly, some comparisons between these mechanisms can be drawn in other SM-producing Enterobacteriaceae (Figure 1a). *Erwinia* is a genus of plant pathogenic bacteria that produce SMs, where Hfq and ArcZ have both been implicated in virulence (34), while HexA is a negative regulator of secondary metabolites in these bacteria (35). Similar parallels can also be seen from *Serratia* (36-38) and *Pseudomonas* (39), two other prolific SM producers. Although further investigations will be required to ascertain whether these apparent similarities represent identical mechanisms, the conserved nature of ArcZ in other SM-producing Enterobacteriaceae could suggest that this strategy may yield fresh avenues for rapid investigation into SM biosynthesis in other taxa.

## ONLINE METHODS

### Bacterial culture conditions

All *Photorhabdus* and *Xenorhabdus* strains were grown in LB with shaking for at least 16 hours at 30°C. *E. coli* strains were grown in LB with shaking for at least 16 hours at 37°C. The medium was supplemented with chloramphenicol (34 µg/ml), ampicillin (100 µg/ml), rifampicin (50µg/ml) or kanamycin (50 µg/ml) when appropriate. Promotor exchange mutants were induced by adding L-arabinose (2%, v/v) to the cultures. All plasmids and strains used in this study are listed in Supplementary Tables 13 & 14.

### Nematode bioassays

All nematodes were cultivated in *Galleria mellonella* and collected on white traps as previously described. Nematode bioassays were also performed as described elsewhere (4).

### Creation of transposon mutant library

For the transposon mutagenesis, the plasmid pSAM\_Kan (containing the mariner transposon) was constructed using pSAM\_BT (40) as a template. To do this, the plasmid was linearized using the primers NN191/NN192. The kanamycin resistance cassette was amplified from the pCOLA\_ara\_tacl plasmid using the primers NN193/NN194 introducing complementary overhangs to pSAM\_BT at both ends of the PCR fragment. The kanamycin resistance cassette was fused with the linearized pSAM\_BT plasmid using Hot Fusion cloning thereby replacing the erythromycin resistance cassette with kanamycin resistance. *E. coli* ST18 was transformed with the plasmid pSAM\_Kan and further used for the creation

of the transposon mutant library of *P. laumondii* TTO1 through conjugation. Transposon-insertion mutants were selected on LB agar containing kanamycin. All primer sequences are listed in Supplementary Table 15.

### **Construction of mutant strains**

For the deletion of the majority of ArcZ in *P. laumondii* TTO1, a 1123 bp upstream and a 1014 bp downstream product was amplified using the primers NN276/NN277 and NN278/NN279, respectively. The PCR products were fused using the complementary overhangs introduced by the primers and cloned into the *Pst*I and *Bgl*II linearized pEB17 plasmid. The resulting plasmid was used for transformation of *E. coli* s17-1  $\lambda$ pir. Conjugation of the plasmid in *P. laumondii* strains and generation of deletion strains by homologous recombination through counter selection was done as previously described (41). Deletion mutants were verified by PCR using the primers NN281/NN282 yielding a 632 bp fragment for mutants genetically equal to the WT and a 502 bp fragment for the desired deletion mutant. Complementation of the ArcZ deletion was achieved by inserting the full and intact version of ArcZ at the original locus. To do this, a 2207 bp PCR product was amplified using the primers NN276/NN279 including the upstream and downstream region required for homologous recombination and the full length ArcZ. The fragment was cloned into pEB17 as described above. The verified plasmid construct was used for transformation of *E. coli* s17-1  $\lambda$ pir cells. The plasmid was transferred into *P. laumondii*  $\Delta$ arcZ by conjugation and integrated into the genome of *P. laumondii*  $\Delta$ arcZ by homologous recombination. The knock-in mutant was generated by a second homologous recombination through counter selection on LB plates containing 6% sucrose. Knock-in mutants were verified by PCR using the primers NN281/NN282 yielding a 632 bp fragment. The same strategy was used for the construction of the mutant strains in *X. szentirmaii*. To generate the promotor exchange mutants in front of *gxpS*, the plasmid pCEPKMR\_ORF00346 was transferred into *X. szentirmaii*  $\Delta$ arcZ and *X. szentirmaii*  $\Delta$ hfq by conjugation and integrated into the genome by homologous recombination.

### **DNA extraction**

Genomic DNA was extracted using the Genra Puregene Yeast/Bact Kit (Qiagen) following the manufacturer's instructions. For sequencing of transposon-insertion mutants, genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen).

### **DNA sequencing and identification of transposon insertion site**

DNA isolated from the transposon-insertion mutants was sequenced on the Illumina NextSeq platform. DNA libraries were constructed using the Nextera XT DNA preparation

kit (Illumina) and whole genome sequencing was performed using 2 x 150bp paired-end chemistry. A sequencing depth of >50× was targeted for each sample. Genomes were assembled with SPAdes (v 3.10.1) (42) and annotated with Prokka v 1.12 (43). Completed genome sequences were analysed and viewed in Geneious v 6 & 9.1 (<https://www.geneious.com>).

### **RNA extraction, sequencing and analysis**

Pre-cultures of *P. laumondii* TTO1, *X. szentirmaii* DSM16338, and their respective ArcZ deletion and knock-in mutants were grown in LB broth overnight with shaking, at 30 °C. The following day, the pre-cultures were used to inoculate fresh LB at an OD<sub>600</sub> of 0.3. Cells were grown to mid-exponential phase (OD values for each experiment can be found in Supplementary Table 3). RNA was extracted using the RNeasy Mini Kit (Qiagen) following the manufacturer's instructions. To facilitate cell lysis, the cells were pelleted and snap frozen in liquid nitrogen for 1 min after removing the supernatant. After thawing and resuspending in lysis buffer, the cells were vortexed for 30 sec before proceeding with the protocol. RNA for small RNA libraries were extracted in duplicate, during the mid-exponential phase for *P. laumondii* TTO1 and *X. szentirmaii*.

RNA was sequenced with 150bp paired-end sequencing by Novogene following rRNA depletion with a RiboZero kit and library preparation following the Illumina protocol for strand-specific libraries. Raw data was trimmed using Trimmomatic (44) and mapped to the reference genome downloaded from NCBI (NC\_005126.1 for *P. laumondii* and NZ\_NIBV00000000.1 for *X. szentirmaii*) using bowtie2 (v2.3.4.3) (45). Resulting .sam files were converted to .bam files using samtools (v1.8) (46) and featureCounts (a part of the subread package) (47) was used to count reads mapping to annotated genes. Count files were then uploaded to degust (<http://degust.erc.monash.edu/>) and analyzed using the voom/limma method of normalization. Only genes with an absolute fold change >2 and false discovery rate < 0.01 were considered significantly regulated. Statistical analysis was performed in R (v 3.6.1) on the degust platform, where exact code is available to view.

### **Northern blot analysis**

For Northern blot analysis, total RNA was prepared and analyzed as described previously (48). Briefly, RNA samples were separated on 6% polyacrylamide / 7 M urea gels and transferred to Hybond-XL membranes (GE Healthcare) by electro-blotting. Membranes were hybridized in Roti-Hybri-Quick buffer (Roth) at 42°C with gene-specific [<sup>32</sup>P] end-labeled DNA oligonucleotides, and washed in three subsequent steps with SSC (5x, 1x, 0.5x) / 0.1% SDS wash buffer. Signals were visualized on a Typhoon FLA 7000

phosphorimager (FUJIFILM). Oligonucleotides for Northern blot analyses are listed in Supplementary Table 15.

### **Compensatory base mutation and GFP fluorescence assay**

Plasmids pMH078 and pMH079 were generated using Gibson assembly (49). For plasmid pMH078 the *arcZ* gene was amplified using *P. laumondii* TTO1 genomic DNA with oligonucleotides KPO-6147 and KPO-6148 and fused into a pEVS143 vector backbone (50), linearized with KPO-0092 and KPO-1397. To construct plasmid pMH079, the 5' UTR and the first 20 aa of *hexA* were amplified using *P. laumondii* TTO1 genomic DNA with KPO-6145 and KPO-6146, and the pXG10-*gfp* vector (51) was linearized with KPO-1702 and KPO-1703. pMH078 and pMH079 served as templates to insert single point mutations in the *arcZ* gene as well as the *hexA* 5' UTR using site-directed mutagenesis and oligonucleotide combinations KPO-6156/KPO-6157 and KPO-6164/KPO-6165, respectively, yielding plasmids pMH080 and pMH081.

Target regulation using GFP reporter fusions was analyzed as described previously (51). *E. coli* Top10 cells were grown overnight in LB medium (37°C, 200 rpm shaking conditions). Three independent cultures were used for each strain. Cells were washed in PBS and GFP fluorescence intensity was determined using a Spark 10 M plate reader (TECAN). Control samples not expressing fluorescence proteins were used to subtract background fluorescence.

### **CappableSeq analysis**

Cappable seq was performed as previously described (52) by Vertis Biotechnologies (Germany). Raw sequences were trimmed with Trimmomatic (44) and mapped with bowtie2 (45) to NC\_005126.1 for *P. laumondii* and NZ\_NIBV00000000.1 for *X. szentirmaii*. Transcriptional start sites were detected using readXplorer's (v2.2.3) (53) built in TSS detection function with the following settings: use only single perfect matches, minimum number of read starts = 100, minimum percent coverage increase = 750, detect previously undescribed transcripts, min. transcript extension = 40, max distance to feature of leaderless transcripts = 5, associate neighbouring TSS within 3bp.

### **RIP-seq analysis**

Overnight cultures of *P. laumondii* TTO1 (WT and Hfq::3xFLAG) were inoculated into fresh LB media in duplicate and grown at 30 °C with shaking at 200 rpm. Bacteria were harvested by centrifugation at 4°C, 4000 rpm for 15 min when cells reached OD<sub>600</sub>=0.5 and OD<sub>600</sub>=5.0. Cell pellets were resuspended in 1 mL lysis buffer (20 mM Tris pH 8.0, 150 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM DTT) and pelleted again by centrifugation (5 min, 11,200 g, 4°C). The

supernatants were discarded and the pellets were snap-frozen in liquid nitrogen. After thawing on ice, cells were resuspended in 800  $\mu$ l lysis buffer and transferred into tubes containing 300  $\mu$ l glass beads to break cells via a Bead Ruptor (150 sec, twice, 2 min break on ice in between). After short centrifugation (15,000 g, 4°C), lysates were transferred into fresh precooled tubes and centrifuged for 30 minutes at 15,200 g at 4°C. The cleared lysates were transferred into fresh tubes and incubated with 35  $\mu$ l FLAG-antibody (Monoclonal ANTI-FLAG M2, Sigma, #F1804) with rotation for 45 min at 4°C, followed by addition of 75  $\mu$ l Protein G Sepharose (Sigma, #P3296) and rotating for 45 min at 4°C again. After five wash steps with lysis buffer (via inverting the tube gently and centrifuging for 4 min at 4°C), samples were subjected to RNA and protein separation by Phenol:Chloroform:Isoamylalcohol (P:C:I; 25:24:1, pH 4.5, Roth) extraction. The upper phase (~ 500  $\mu$ l) was transferred into a fresh tube and precipitated overnight at -20°C with 1.5 ml EtOH:Na(acetate) (30:1) and 1.5  $\mu$ l GlycoBlue (#AM9516, Ambion). After centrifugation for 30 minutes at 11,200 rpm at 4°C, RNA pellets were washed with 500  $\mu$ l 70% EtOH, dried and resuspended in 15.5  $\mu$ l nuclease-free H<sub>2</sub>O. RNA was treated with 2  $\mu$ l DNase I, 0.5  $\mu$ l RNase inhibitor and 2  $\mu$ l 10x DNase buffer at 37°C for 30 min. Afterwards, samples were supplemented with 100  $\mu$ l H<sub>2</sub>O, and again subjected to P:C:I extraction. The upper phase (~120  $\mu$ l) was transferred into a fresh tube with addition of 2.5-3 volumes (~350  $\mu$ l) of EtOH:Na(acetate) (30:1) and stored at -20°C overnight for RNA precipitation. RNA pellets were harvested via centrifugation for 30 min at 13,000 rpm, 4°C, and washed with 500  $\mu$ l 70% EtOH, dried and resuspended in nuclease-free H<sub>2</sub>O. cDNA libraries were prepared using the NEBNext Small RNA Library Prep Set for Illumina (NEB, #E7300S) according to the manufacturer's instructions and sequenced on a HiSeq 1500 system in single-read mode with 100 nt read length.

For RIP-seq analysis, the enriched/control sample pairs were normalized by the number of raw reads present after trimming. Depth counts of all samples were obtained using samtools (v1.8) (46). Only nucleotide positions with a depth of at least 50 reads in the enriched samples were taken for further analysis. The corresponding depth in the unenriched samples was matched for each nucleotide. A region was considered to be enriched if the enrichment factor was at least three and the corresponding 'enriched' nucleotide was present in both sample pairs. Finally, we considered a region to be enriched if more than five consecutive nucleotides were identified as enriched.

### **ArcZ binding prediction**

ArcZ from *E. coli* was used to define the boundaries of ArcZ in *Xenorhabdus* and *Photorhabdus*. We then took our annotated ArcZ sequence together with several ArcZ homologs from other Enterobacteriaceae (listed in Supplementary Table 9) and used the

online CopraRNA tool (23), a part of the Freiburg RNA tools suite (54), with default parameters.

### **Metabolite extraction and HPLC-MS/MS analysis**

Fresh 10 ml of LB was inoculated with an overnight culture to an  $OD_{600} = 0.1$ . After 72 h of cultivation at 30°C with shaking, 1 ml of the culture was removed from the culture, centrifuged for 20 min at 13,300 rpm and the supernatant was directly subjected for HPLC-UV/MS analysis using a Dionex Ultimate 3000 system with a Bruker AmaZon X mass spectrometer. The compounds peak areas were quantified using TargetAnalysis 1.3 (Bruker). All analyzed compounds are listed in Supplementary Table 16.

### **Proteome analysis**

The detail of the proteomics procedure was previously published (55). In short, to extract proteins from *P. laumondii*, frozen cell pellets were resuspended in 300  $\mu$ L lysis buffer (0.5% Na-desoxycholate in 100 mM  $NH_4HCO_3$ ), and incubated at 95°C for 10 min. The protein concentration in the supernatant was determined with a BCA Protein Assay Kit (Thermo Fisher, #23252). Reduction and alkylation was performed at 95 °C using 5mM TCEP and 10mM Chloroacetamide for 15 min. 50  $\mu$ g of protein was transferred to fresh reaction tubes and protein digestion was carried out overnight at 30 °C with 1  $\mu$ g trypsin (Promega). After digest, the peptides were desalted using CHROMABOND Spin columns (Macherey-Nagel) that were conditioned with 500  $\mu$ L of acetonitrile and equilibrated with 500  $\mu$ L and 150  $\mu$ L 0.1% TFA. After loading the peptides were washed with 500  $\mu$ L 0.1% TFA in 5:95 acetonitrile:water, peptides were eluted with 400  $\mu$ L 0.1% TFA in 50:50 acetonitrile:water. Peptides were concentrated and dried under vacuum at 50°C and dissolved in 100  $\mu$ L 0.1% TFA by 25 s of sonication and incubation at 22°C under shaking at 1200 rpm for 5 min. 1  $\mu$ g peptide was analyzed using liquid chromatography-mass spectrometry (LC-MS/MS).

The LC-MS/MS analysis including label-free quantification was carried out as previously described (55) , with minor modifications. LC-MS/MS analysis of protein digests was performed on Q-Exactive Plus mass spectrometer connected to an electrospray ion source (Thermo Fisher Scientific). Peptide separation was carried out using Ultimate 3000 nanoLC-system (Thermo Fisher Scientific), equipped with packed in-house C18 resin column (Magic C18 AQ 2.4  $\mu$ m, Dr. Maisch). The peptides were first loaded onto a C18 precolumn (preconcentration set-up) and then eluted in backflush mode with a gradient from 98 % solvent A (0.15 % formic acid) and 2 % solvent B (99.85 % acetonitrile, 0.15 % formic acid) to 35 % solvent B over 30 min. Label-free quantification was done using Progenesis QI software (Nonlinear Dynamics, v2.0), MS/MS search was performed in MASCOT (v2.5, Matrix Science) against the Uniprot *Photorhabdus laumondii* protein database. The following

search parameters were used: full tryptic search with two missed cleavage sites, 10ppm MS1 and 0.02 Da fragment ion tolerance. Carbamidomethylation (C) as fixed, oxidation (M) and deamidation (N,Q) as variable modification. Progenesis outputs were further processed with SafeQuant (56).

### **DATA AVAILABILITY**

All .mzXML files from HPLC-MS runs are available at MassIVE (<https://massive.ucsd.edu>) under the ID MSV000084163. Raw sequence data is available at the European nucleotide archive (<https://www.ebi.ac.uk/ena/>) under project accession numbers PRJEB33827 and PRJEB24159. The proteomic data can be accessed at PRIDE (<https://www.ebi.ac.uk/pride/>) with the project accession number PXD019095. Source Data are provided with this paper.

### **CORRESPONDING AUTHOR**

Correspondence to Helge B. Bode.

### **ACKNOWLEDGEMENTS**

This work was funded in part by the DFG (SFB 902, Project B17) and the LOEWE Centre for Translational Biodiversity Genomics (TBG) supported by the State of Hesse. K.P. acknowledges funding by DFG (EXC 2051, Project 390713860), the Vallee Foundation, and the European Research Council (StG-758212). We thank Andrew Goodman for providing pSAM-BT and helpful discussions. We thank Laura Pöschel and Antje K. Heinrich for plasmid construction.

### **CONTRIBUTIONS**

These authors contributed equally: Nick Neubacher, Nicholas J. Tobias, Michaela Huber. N.N., N.J.T., M.H., X.C., A.L., T.G. and S.J.P. performed experiments except sequencing of the transposon insertion mutants, which was performed by S.J.P., and T.P.S. N.N., N.J.T., M.H., K.P. and H.B.B. designed the study, discussed the results and commented on the manuscript. N.N., N.J.T. and M.H. analyzed and interpreted the data. N.N., N.J.T. and M.H. wrote the manuscript. All authors read and approved the final manuscript.

### **COMPETING INTERESTS**

The authors declare no competing interests.

## FIGURES AND LEGENDS

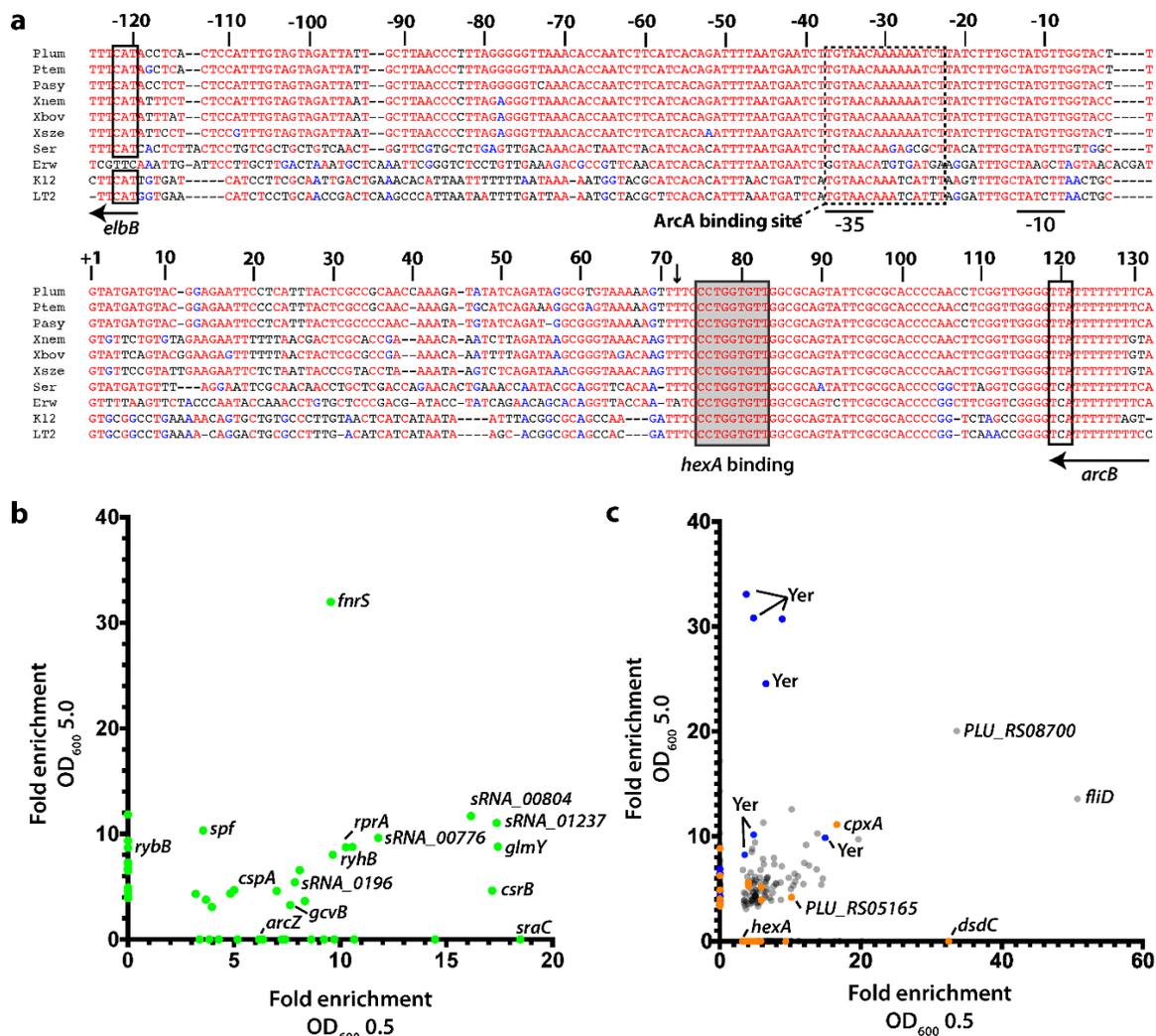
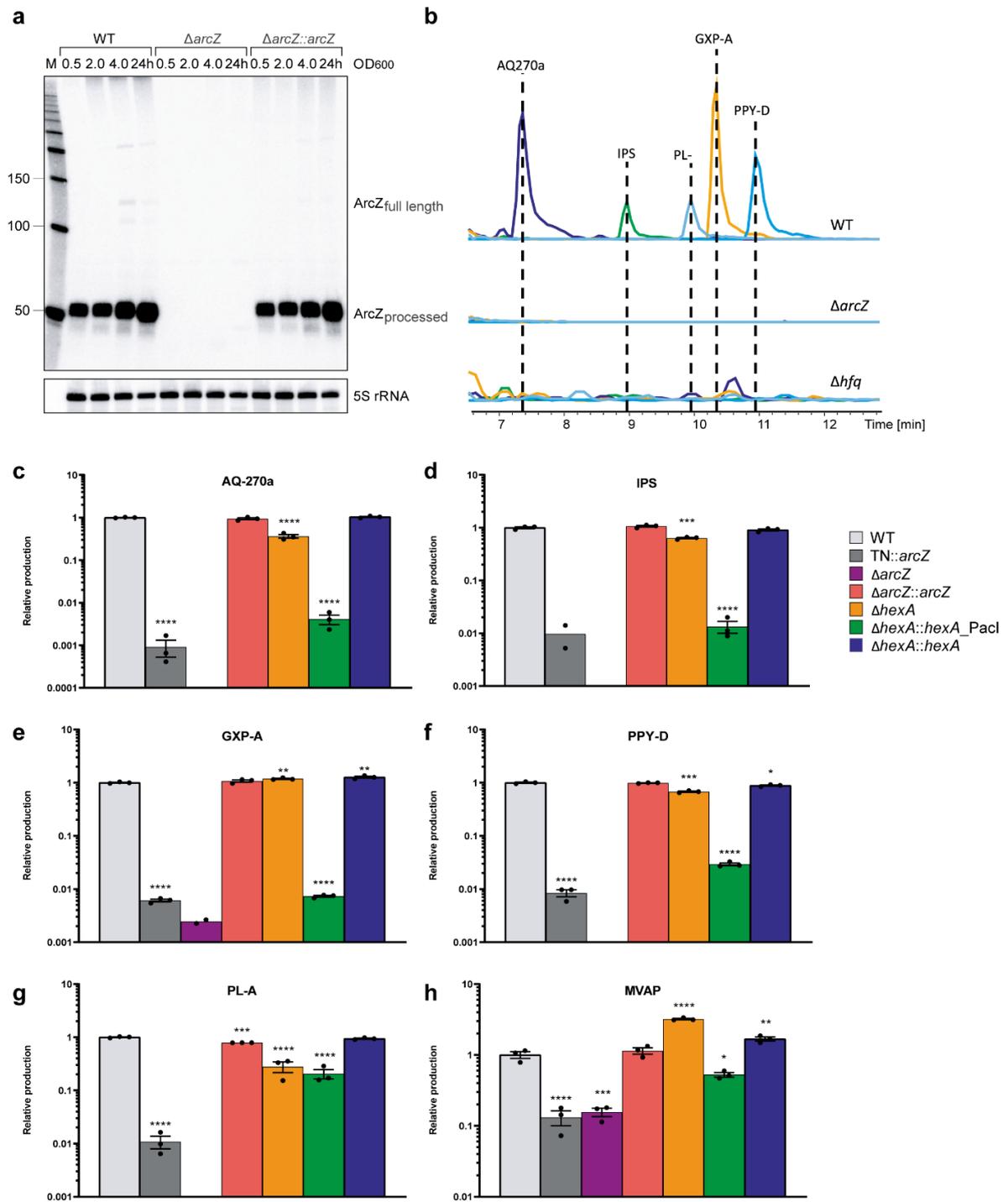


Figure 1

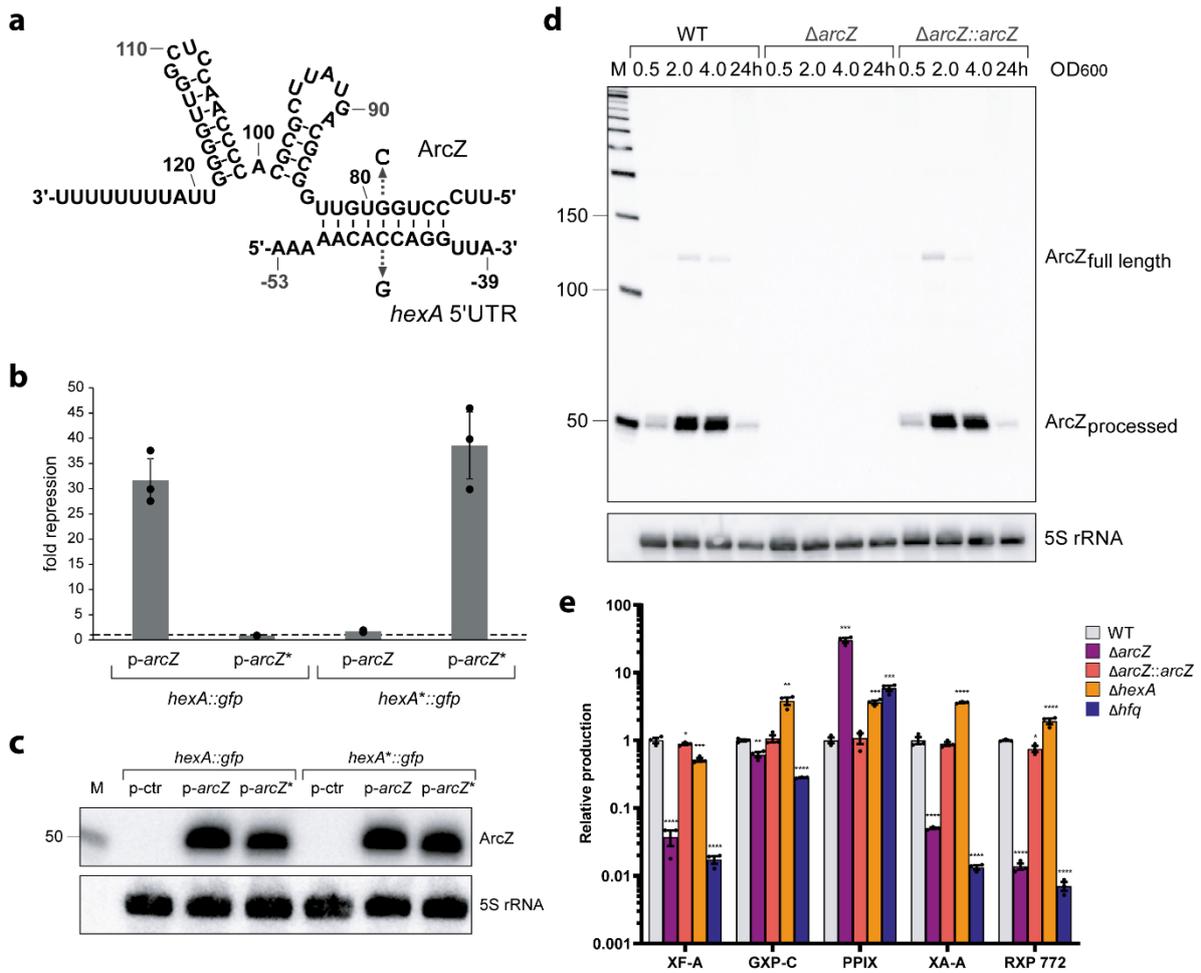
**a.** Alignment of *arcZ* sequences from *P. laumondii* TTO1, *P. temperata*, *P. asymbiotica*, *X. nematophila*, *X. bovienii*, *X. szentirmaii*, *Serratia marcescens*, *Erwinia amylovora*, *E. coli* K12 and *Salmonella typhimurium* LT2. Numbers refer to *P. laumondii* sequence. The +1 indicates the transcriptional start of the 129 nt *arcZ* sequence. Indicated are the start codon of *elbB* and the stop codon of *arcB*, -10 and -35 binding regions, as well as the conserved ArcA binding region (11) and the region of base-pairing to *hexA*. The site of ArcZ cleavage is indicated by an arrow. **b.** RIP-seq enrichment in regions of sRNAs and **c.** mRNAs in a strain containing Hfq::3xFLAG when compared to the untagged control strain at both optical densities. For a complete list of enriched regions see Supplementary Tables 4 and 5. Blue dots represent SM-related mRNAs, while orange dots represent mRNAs associated with annotated regulators.



**Figure 2**

**a.** ArcZ expression in *P. laumondii* WT,  $\Delta arcZ$  and  $\Delta arcZ::arcZ$  cells detected by Northern blot analysis. Total RNA samples were collected at three different OD<sub>600</sub> values (0.5; 2 and 4) and after 24 h of growth. Probing for 5S rRNA served as loading control. Representative blot image of two biologically independent replicates. **b.** Comparison of relative production titers of the major SMs produced by *P. laumondii* WT,  $\Delta arcZ$  and  $\Delta hfq$ . Depicted are the extracted ion chromatograms of anthraquinone (AQ-270a), isopropylstilbene (IPS),

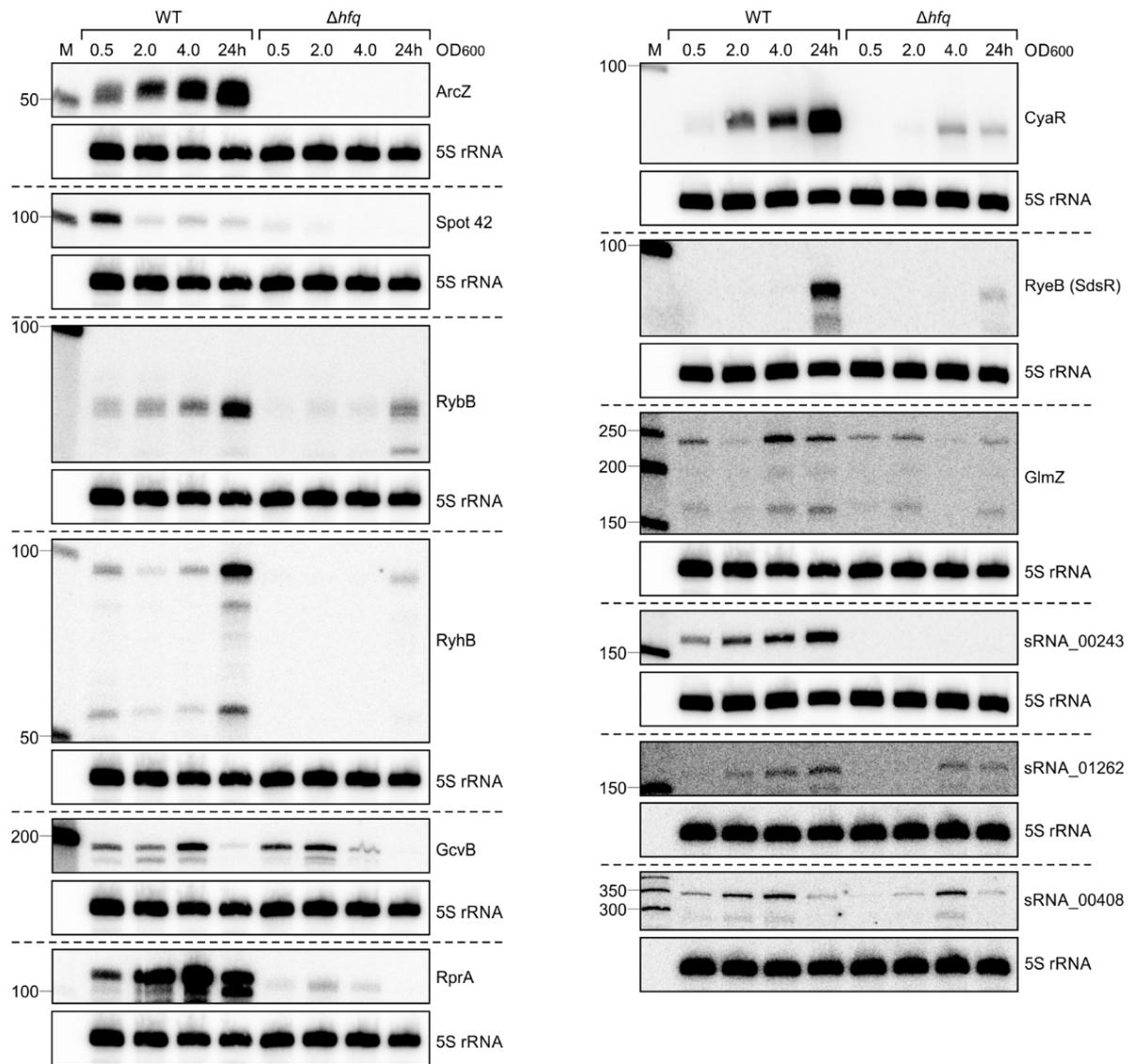
phurealipid A (PL-A), GameXPeptide A (GXP-A) and photopyrone D (PPY-D) in the WT and the mutant strains. **c-h.** HPLC-MS quantification of **c.** AQ-270a, **d.** IPS, **e.** GXP-A, **f.** PPY-D, **g.** PL-A and **h.** MVAP in *P. laumondii* WT (light grey), TN::*arcZ* (grey),  $\Delta$ *arcZ* (purple),  $\Delta$ *arcZ*::*arcZ* (red),  $\Delta$ *hexA* (orange),  $\Delta$ *hexA*::*hexA*\_Pacl (green) and  $\Delta$ *hexA*::*hexA* (blue). All bars represent relative production in comparison to the wild type. Data are presented as mean values +/- SEM. Dots represent biologically independent replicates (n=3). Asterisks indicate statistical significance (\* p<0.05, \*\* p<0.005, \*\*\* p<0.0005, \*\*\*\* p<0.0001) of relative production compared to WT production levels. For panels c to h statistical significances were calculated using a two-sided unpaired t-test. Exact p values (left to right, respectively) correspond to p= n.d. (not determined), n.d., 0.24, 0.0005, <0.0001, 0.12 in panel c; <0.0001, n.d., 0.20, <0.0001, <0.0001, 0.24 in panel d; <0.0001, n.d., 0.18, 0.0018, <0.0001, 0.0037 in panel e; <0.0001, n.d., 0.56, 0.0003, <0.0001, 0.013 in panel f; <0.0001, n.d., 0.0001, <0.0001, <0.0001, 0.059 in panel g; 0.0006, 0.0008, 0.34, <0.0001, 0.011, 0.0058 in panel g. Details of all analyzed compounds can be found in Supplementary Table 16.

**Figure 3**

**a.** Predicted base-pairing interaction of ArcZ with the 5' UTR of the *hexA* mRNA. Arrows indicate the single nucleotide mutations tested in **b**. **b.** Measurement of GFP signals derived from co-expression of a plasmid harbouring the 5' UTR of *hexA* fused to *gfp* (*hexA::gfp*) or the same fusion with a single point mutation (C-46G, *hexA\*::gfp*) with p-ctr, p-arcZ or p-arcZ\* (G79C). GFP levels of strains carrying the control plasmid (p-ctr) were set to 1. Data are presented as mean values +/- SD. Dots represent biologically independent replicates (n=3). **c.** Northern blot analysis of ArcZ expression corresponding to the GFP expression assay shown in **b**. 5S rRNA served as loading control. **d.** ArcZ expression in *X. szentirmaii* WT,  $\Delta arcZ$  and  $\Delta arcZ::arcZ$  cells detected by Northern blot analysis. Total RNA samples were collected at three different OD<sub>600</sub> values (0.5; 2 and 4) and after 24 h of growth. 5S rRNA served as loading control. **e.** HPLC-MS quantification of strains of *X. szentirmaii* WT (light grey),  $\Delta arcZ$  (purple),  $\Delta arcZ::arcZ$  (red),  $\Delta hexA$  (orange) and  $\Delta hfq$  (blue). Shown is the relative production of xenofuranone A (XF-A), GameXPepide C (GXP-C), protoporphyrin IX (PPIX), xenoamicin A (XA-A) and rhabdopeptide 772 (RXP 772). See also Supplementary Table 16. Data are presented as mean values +/- SEM. Dots represent biologically

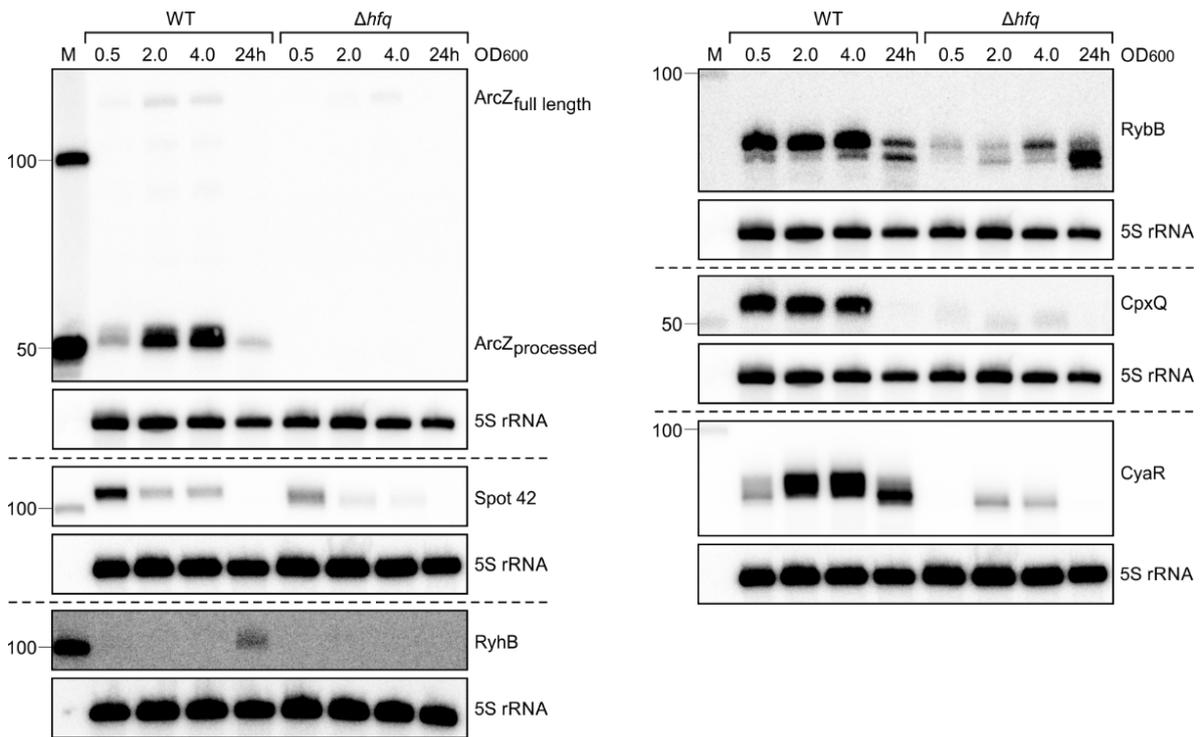
independent replicates (n=3). Asterisks indicate statistical significance (\*  $p < 0.05$ , \*\*  $p < 0.005$ , \*\*\*  $p < 0.0005$ , \*\*\*\*  $p < 0.0001$ ) of relative production compared to WT production levels. For panel e statistical significances were calculated using a two-sided unpaired t-test. Exact p values (left to right, respectively) correspond to  $p = < 0.0001$ , 0.021, 0.0001,  $< 0.0001$  for XF-A, 0.0034, 0.059, 0.0021,  $< 0.0001$  for GXP-C, 0.0001, 0.8, 0.0008, 0.0007 for PPIX,  $< 0.0001$ , 0.14,  $< 0.0001$ ,  $< 0.0001$  for XA-A and  $< 0.0001$ , 0.02, 0.0041,  $< 0.0001$  for RXP 772. Details of all analyzed compounds can be found in Supplementary Table 16.





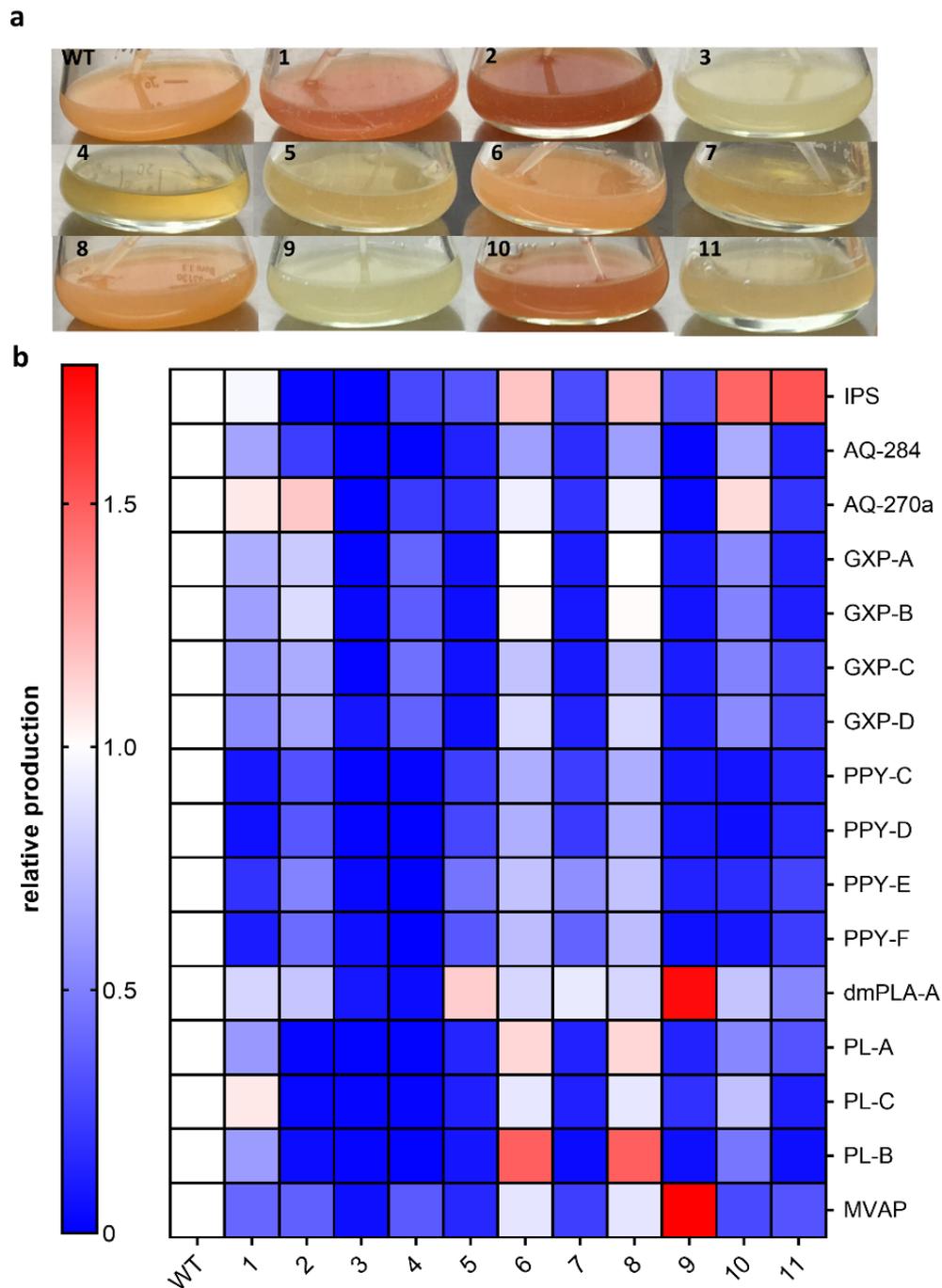
### Extended Data Figure 1

Expression of various sRNAs in *P. laumondii* at different time points. RNA samples of *P. laumondii* WT and  $\Delta hfg$  strains were taken at three different OD<sub>600</sub> values (0.5, 2 and 4) and after 24 h of growth. The RNA was loaded on Northern blots and probed for the indicated sRNAs. Probing for 5S rRNA served as loading control.



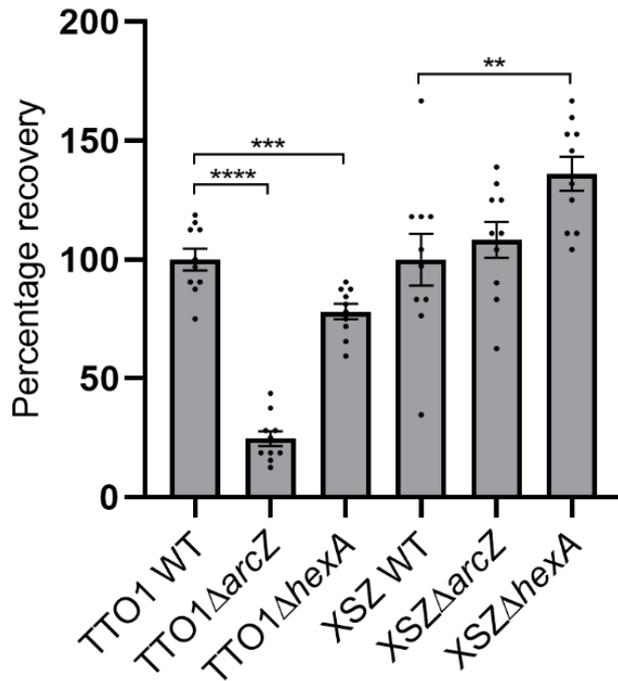
**Extended Data Figure 2**

Expression of various sRNAs in *X. szentirmaii* at different time points. RNA samples of *X. szentirmaii* WT and  $\Delta hfq$  strains were taken at three different OD<sub>600</sub> values (0.5, 2 and 4) and after 24 h of growth. The RNA was loaded on Northern blots and probed for the indicated sRNAs. Probing for 5S rRNA served as loading control.



### Extended Data Figure 3

Phenotype of transposon insertion mutants of *P. laumondii*. **a.** Differences in pigmentation of transposon insertion mutant liquid cultures compared to WT. Depicted are eleven transposon insertion mutants and a WT culture after 3 d of cultivation at 30°C with shaking. **b.** SM-profiles of the transposon insertion mutants. Relative SM production was quantified from duplicates using TargetAnalysis (Bruker) and compared to the WT of *P. laumondii* after 72 h cultivation at 30°C with shaking. Mutant 3 was analyzed further and the transposon insertion was identified in the *arcZ* gene.



#### Extended Data Figure 4

Infective juvenile development to hermaphrodites with strains of *P. laumondii* and *X. szentirmaii*. Data are presented as mean values  $\pm$  SEM. Dots represent biologically independent replicates ( $n=10$ ). Asterisks indicate statistical significance (\* $p<0.05$ , \*\* $p<0.005$ , \*\*\* $p<0.0005$ , \*\*\*\* $p<0.00005$ ) of relative recovery compared to WT recovery levels. Statistical significances were calculated using a two-sided unpaired test. Exact  $p$  values (left to right, respectively) for *P. laumondii* TTO1 correspond to  $p= <0.0001$ ,  $0.0006$  and for *X. szentirmaii* to  $p= 0.56$ ,  $0.0094$ .

## REFERENCES

1. Stock SP, Campbell JF, & Nadler SA (2001) Phylogeny of *Steinernema travassos*, 1927 (*Cephalobina: Steinernematidae*) inferred from ribosomal DNA sequences and morphological characters. *J Parasitol* 87(4):877-889.
2. Forst S, Dowds B, Boemare N, & Stackebrandt E (1997) *Xenorhabdus* and *Photorhabdus* spp.: bugs that kill bugs. *Annu Rev Microbiol* 51:47-72.
3. Engel Y, Windhorst C, Lu X, Goodrich-Blair H, & Bode HB (2017) The Global Regulators Lrp, LeuO, and HexA Control Secondary Metabolism in Entomopathogenic Bacteria. *Front Microbiol* 8:209.
4. Tobias NJ, *et al.* (2017) *Photorhabdus*-nematode symbiosis is dependent on hfq-mediated regulation of secondary metabolites. *Environ Microbiol* 19(1):119-129.
5. Santiago-Frangos A & Woodson SA (2018) Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip Rev RNA* 9(4):e1475.
6. Holmqvist E & Vogel J (2018) RNA-binding proteins in bacteria. *Nat Rev Microbiol* 16(10):601-615.
7. Carrier MC, Lalaouna D, & Masse E (2018) Broadening the Definition of Bacterial Small RNAs: Characteristics and Mechanisms of Action. *Annu Rev Microbiol* 72:141-161.
8. Vogel J & Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* 9(8):578-589.
9. Chao Y, Papenfort K, Reinhardt R, Sharma CM, & Vogel J (2012) An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* 31(20):4005-4019.
10. Papenfort K, *et al.* (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol Microbiol* 74(1):139-158.
11. Mandin P & Gottesman S (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO J* 29(18):3094-3107.
12. Chao Y, *et al.* (2017) In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Mol Cell* 65(1):39-51.
13. Beemelmanns C, Guo H, Rischer M, & Poulsen M (2016) Natural products from microbes associated with insects. *Beilstein J Org Chem* 12:314-327.
14. Shi YM & Bode HB (2018) Chemical language and warfare of bacterial natural products in bacteria-nematode-insect interactions. *Nat Prod Rep* 35(4):309-335.
15. Brachmann AO, *et al.* (2013) Pyrones as bacterial signaling molecules. *Nat Chem Biol* 9(9):573-578.
16. Brameyer S, Kresovic D, Bode HB, & Heermann R (2015) Dialkylresorcinols as bacterial signaling molecules. *Proc Natl Acad Sci U S A* 112(2):572-577.
17. Joyce SA, *et al.* (2008) Bacterial biosynthesis of a multipotent stilbene. *Angew Chem Int Ed Engl* 47(10):1942-1945.
18. Cai X, *et al.* (2017) Entomopathogenic bacteria use multiple mechanisms for bioactive peptide library design. *Nat Chem* 9(4):379-386.
19. Crawford JM, Portmann C, Zhang X, Roeffaers MB, & Clardy J (2012) Small molecule perimeter defense in entomopathogenic bacteria. *Proc Natl Acad Sci U S A* 109(27):10821-10826.
20. Theodore CM, King JB, You J, & Cichewicz RH (2012) Production of cytotoxic glidobactins/luminmycins by *Photorhabdus asymbiotica* in liquid media and live crickets. *J Nat Prod* 75(11):2007-2011.
21. Lango-Scholey L, Brachmann AO, Bode HB, & Clarke DJ (2013) The expression of *stlA* in *Photorhabdus luminescens* is controlled by nutrient limitation. *PLoS One* 8(11):e82152.

22. Kontnik R, Crawford JM, & Clardy J (2010) Exploiting a global regulator for small molecule discovery in *Photobacterium luminescens*. *ACS Chem Biol* 5(7):659-665.
23. Wright PR, *et al.* (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res* 42:W119-123.
24. Joyce SA & Clarke DJ (2003) A *hexA* homologue from *Photobacterium* regulates pathogenicity, symbiosis and phenotypic variation. *Mol Microbiol* 47(5):1445-1457.
25. Bouvier M, Sharma CM, Mika F, Nierhaus KH, & Vogel J (2008) Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol Cell* 32(6):827-837.
26. Sharma CM, Darfeuille F, Plantinga TH, & Vogel J (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* 21(21):2804-2817.
27. Yang Q, Figueroa-Bossi N, & Bossi L (2014) Translation enhancing ACA motifs and their silencing by a bacterial small regulatory RNA. *PLoS Genet* 10(1):e1004026.
28. Wenger RH (2000) Mammalian oxygen sensing, signalling and gene regulation. *J Exp Biol* 203(Pt 8):1253-1263.
29. Chen JJ & London IM (1995) Regulation of protein synthesis by heme-regulated eIF-2 alpha kinase. *Trends Biochem Sci* 20(3):105-108.
30. Qi Z, Hamza I, & O'Brian MR (1999) Heme is an effector molecule for iron-dependent degradation of the bacterial iron response regulator (Irr) protein. *Proc Natl Acad Sci U S A* 96(23):13056-13061.
31. Rao AU, Carta LK, Lesuisse E, & Hamza I (2005) Lack of heme synthesis in a free-living eukaryote. *Proc Natl Acad Sci U S A* 102(12):4270-4275.
32. Bolla R (1979) Developmental nutrition of nematodes: the biochemical role of sterols, heme compounds, and lysosomal enzymes. *J Nematol* 11(3):250-259.
33. Bode E, *et al.* (2019) Promoter Activation in Deltahfq Mutants as an Efficient Tool for Specialized Metabolite Production Enabling Direct Bioactivity Testing. *Angew Chem Int Ed Engl* 58(52):18957-18963.
34. Zeng Q, McNally RR, & Sundin GW (2013) Global small RNA chaperone Hfq and regulatory small RNAs are important virulence regulators in *Erwinia amylovora*. *J Bacteriol* 195(8):1706-1717.
35. Mukherjee A, Cui Y, Ma W, Liu Y, & Chatterjee AK (2000) *hexA* of *Erwinia carotovora* ssp. *carotovora* strain Ecc71 negatively regulates production of RpoS and *rsmB* RNA, a global regulator of extracellular proteins, plant virulence and the quorum-sensing signal, N-(3-oxohexanoyl)-L-homoserine lactone. *Environ Microbiol* 2(2):203-215.
36. Matilla MA, Leeper FJ, & Salmond GP (2015) Biosynthesis of the antifungal haterumalide, oocydin A, in *Serratia*, and its regulation by quorum sensing, RpoS and Hfq. *Environ Microbiol* 17(8):2993-3008.
37. Wilf NM & Salmond GPC (2012) The stationary phase sigma factor, RpoS, regulates the production of a carbapenem antibiotic, a bioactive prodigiosin and virulence in the enterobacterial pathogen *Serratia* sp. ATCC 39006. *Microbiology (Reading)* 158(Pt 3):648-658.
38. Wilf NM, *et al.* (2011) The RNA chaperone, Hfq, controls two luxR-type regulators and plays a key role in pathogenesis and production of antibiotics in *Serratia* sp. ATCC 39006. *Environ Microbiol* 13(10):2649-2666.
39. Shanks RMQ, *et al.* (2017) Suppressor analysis of *eepR* mutant defects reveals coordinate regulation of secondary metabolites and serralyisin biosynthesis by EepR and HexS. *Microbiology (Reading)* 163(2):280-288.
40. Goodman AL, *et al.* (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6(3):279-289.
41. Brachmann AO, *et al.* (2007) A type II polyketide synthase is responsible for anthraquinone biosynthesis in *Photobacterium luminescens*. *Chembiochem* 8(14):1721-1728.

42. Bankevich A, *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455-477.
43. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068-2069.
44. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
45. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359.
46. Li H, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
47. Liao Y, Smyth GK, & Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923-930.
48. Frohlich KS, Haneke K, Papenfort K, & Vogel J (2016) The target spectrum of SdsR small RNA in *Salmonella*. *Nucleic Acids Res* 44(21):10406-10422.
49. Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res* 37(20):6984-6990.
50. Dunn AK, Millikan DS, Adin DM, Bose JL, & Stabb EV (2006) New rfp- and pES213-derived tools for analyzing symbiotic *Vibrio fischeri* reveal patterns of infection and lux expression in situ. *Appl Environ Microbiol* 72(1):802-810.
51. Corcoran CP, *et al.* (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* 84(3):428-445.
52. Tobias NJ, Linck A, & Bode HB (2018) Natural Product Diversification Mediated by Alternative Transcriptional Starting. *Angew Chem Int Ed Engl* 57(20):5699-5702.
53. Hilker R, *et al.* (2016) ReadXplorer 2-detailed read mapping analysis and visualization from one single source. *Bioinformatics* 32(24):3702-3708.
54. Smith C, Heyne S, Richter AS, Will S, & Backofen R (2010) Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. *Nucleic Acids Res* 38:W373-377.
55. Hakobyan A, Liesack W, & Glatter T (2018) Crude-MS Strategy for in-Depth Proteome Analysis of the Methane-Oxidizing Methylocystis sp. strain SC2. *J Proteome Res* 17(9):3086-3103.
56. Glatter T, *et al.* (2012) Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res* 11(11):5145-5156.
57. Tobias NJ, *et al.* (2019) Cyclo(tetrahydroxybutyrate) production is sufficient to distinguish between *Xenorhabdus* and *Photorhabdus* isolates in Thailand. *Environ Microbiol* 21(8):2921-2932.



**SUPPLEMENTARY INFORMATION**

**Supplementary Figure 1.** Sequence of region in *P. laumondii* TTO1 containing predicted *arcZ* sequence

**Supplementary Figure 2.** The ArcZ and Hfq regulon in *P. laumondii*

**Supplementary Figure 3.** RIP-seq enrichment around the region of *hexA*

**Supplementary Figure 4.** 5' UTR of *hexA* including predicted ArcZ binding site

**Supplementary Figure 5.** Alignment of the *hexA* 5' UTR from different species

**Supplementary Notes**

**Supplementary Note 1. Hfq is involved in SM biosynthesis in *Xenorhabdus*.** To confirm if Hfq is also involved in SM regulation in *Xenorhabdus*, we created a knockout of *hfq* in *X. szentirmaii* DSM16338 and performed HPLCMS/MS and RNAseq on the confirmed deletion strains (Supplementary Table 1). In contrast to *Photorhabdus*, the *X. szentirmaii*  $\Delta hfq$  strain only revealed 312 coding sequences significantly regulated compared to the wild type at mid-exponential phase. In accordance with our hypothesis, *hexA* was significantly upregulated (8.4x, FDR<0.01, Supplementary Table 1). Consistent with this observation, the production of nearly all known SMs were decreased (Figure 3e), suggesting a conserved mode of action in *Xenorhabdus*.

**Supplementary Note 2. Identification of sRNAs in *Photorhabdus* and *Xenorhabdus*.** Only very little is known about sRNAs from entomopathogenic bacteria. To identify potential Hfq-binding sRNAs and consequently the Hfq-based regulation of SMs in general, we sequenced the RNA of *P. laumondii* (formerly *P. luminescens*) and *X. szentirmaii* using a library preparation protocol specific for sRNAs. Sequences of the sRNAs from two libraries from each of *Photorhabdus* and *Xenorhabdus* yielded a total of 26,784,563 (13,204,857 and 13,579,706) and 28,813,442 (13,472,683 and 15,340,759) raw reads, respectively. Additionally, we prepared samples from *X. szentirmaii* for CappableSeq, a protocol that differentiates between primary and secondary transcripts (1). We recently reported a data set from *P. laumondii*, which identified 15,500 primary and 3,741 secondary transcripts (2). Here, we reanalyzed these data using stricter cutoff criteria (see Methods) resulting in a total of 6,174 TSSs. The *X. szentirmaii* CappableSeq data led to the identification of 2,196 TSSs (Supplementary Table 2).

By combining data from the CappableSeq experiments data along with RNAseq data from  $\Delta hfq$  and wild type strains (also  $\Delta hfq\Delta hexA$  and  $\Delta hfq::hfq$  in *Photorhabdus* from our previous study (3)), we were able to annotate putative transcripts, 5' untranslated regions (UTRs), 3' UTRs and sRNAs using ANNOgesic (4) (Supplementary Table 17 and 18). The annotated sRNAs were added to those described in the Bacterial sRNA Database (BSRD) (5) yielding a total of 280 and 130 candidates for sRNAs in *Photorhabdus* and *Xenorhabdus*, respectively (Supplementary Tables 17 & 18).

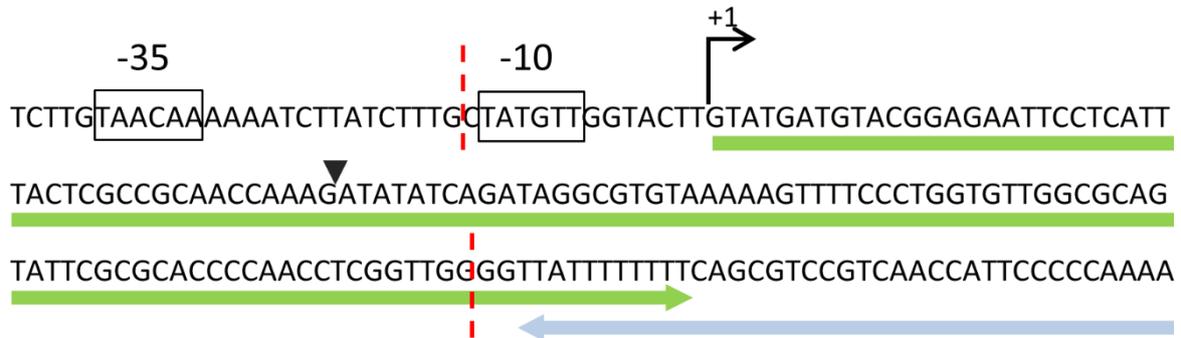
**Supplementary Note 3. Transposon mutant library screen.** A transposon mutant library was constructed to identify genes defective in SM production. Many of the analyzed mutant strains showed severely reduced SM production titers in comparison to the WT strain. In most cases, multiple SM classes were affected by the transposon insertion (Extended Data

Figure 3). On rare occasions, the transposon insertion led to an increase in production of certain SMs. For example, dmPLA-A and MVAP levels were elevated in mutant strain 9 and IPS titers were slightly raised in the TN-mutant strains 10 and 11. Interestingly, the remaining SMs were negatively affected in those strains. As the growth appeared to be affected by the transposon insertion (Supplementary Table 6), it remains uncertain how the growth defects correlate with SM production. For further analysis, we decided to focus on strain 3 that showed only moderate growth defects while at the same time producing reduced SM titers, consistent with the phenotype of the *hfq* deletion mutant.

**Supplementary Note 4. The ArcZ regulon in *Photothabdus* and *Xenorhabdus*.** Since there is a clear overlap between the regulons and functions of Hfq and ArcZ, we performed RNAseq on the  $\Delta arcZ$  strains of *P. laumondii* and *X. szentirmaii*, as well as on their respective knock-in complementation mutants. RNAseq analysis on the deletion of *arcZ* in *Photothabdus* revealed an even broader effect than in our  $\Delta hfq$  mutant, significantly affecting the transcriptional level of 735 coding sequences in *P. laumondii* (FDR<0.01; log<sub>2</sub> fold change >2, Figure 4a, Supplementary Tables 7 & 8). In *X. szentirmaii*, a global effect of the *arcZ* deletion was also observed, albeit only 191 genes were affected in this strain (Supplementary Table 12). In both deletion strains, however, the majority of affected coding sequences were downregulated (Figure 4a & b, Supplementary Tables 8 & 12). In an attempt to identify broader effects, we grouped all the genes that were significantly changed into eight different categories based on their known or proposed function: SM, regulators, virulence, phage related, cell wall, cell processes, hypothetical proteins and unknown. We first included only those genes that were significantly regulated in the *arcZ* deletion mutant and not in the *hfq* deletion mutant (Supplementary Figure 2a). In all cases (except for virulence related and unknown) a clear trend towards downregulation of the transcriptional level could be observed in the deletion of *arcZ*. This trend was also observed in the *hfq* deletion mutant, although somewhat weakened compared to the *arcZ* deletion strain. The knock-in complementation restored the vast majority of observed changes back to WT level (Supplementary Figure 2a). Finally, we looked at genes whose expression was significantly altered in both the *arcZ* and *hfq* deletion strain. The individual categories clustered very closely together as indicated by the median (Supplementary Figure 2b).

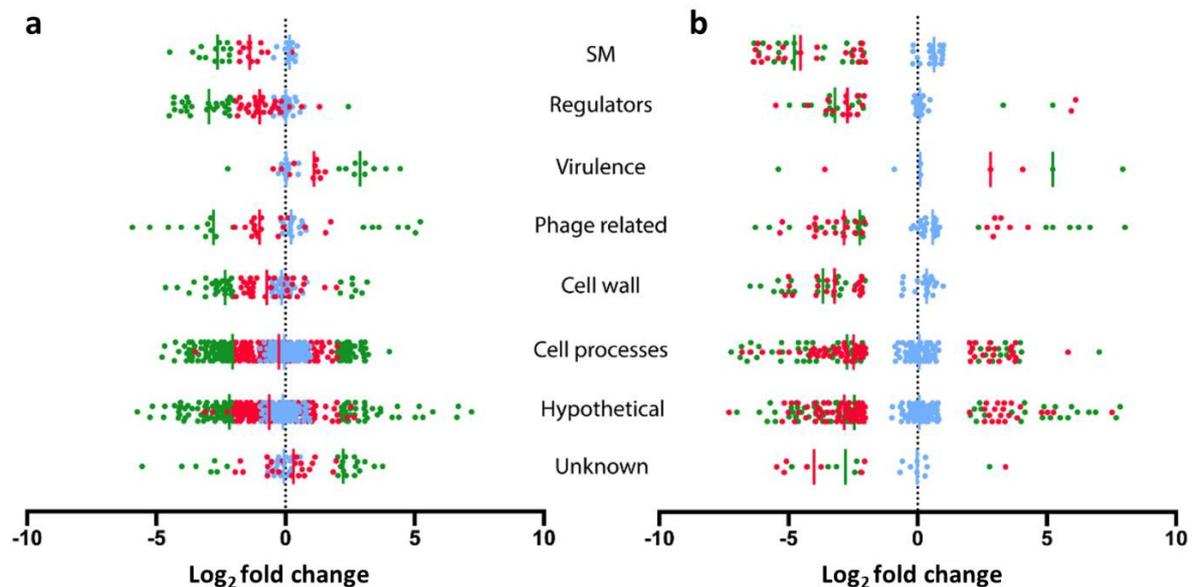
**Supplementary Note 5. Effect of *arcZ* deletion in *Xenorhabdus*.** The drastic reduction in SMs in the deletion mutant was restored with a knock-in complementation of *arcZ* (Figure 3e). We also observed that protoporphyrin IX (PPIX), the direct precursor for heme, was highly overproduced (~30-fold) in the  $\Delta arcZ$  strain of *X. szentirmaii* compared to the WT, suggesting that the regulatory functions of ArcZ in *Photothabdus* and *Xenorhabdus*

possibly go beyond SM production. Since heme is reported to play an important role in nematode growth and development, we used the deletion mutants and complemented strains and performed nematode development assays. Both the WT and  $\Delta arcZ$  strain of *X. szentirmaii* were able to support nematode development after 4 days of inoculation. However, the  $\Delta arcZ$  strain of *P. laumondii* showed a significantly reduced capability to support nematode development (Extended Data Figure 4), consistent with our data showing that isopropylstilbene falls under the Hfq-ArcZ regulatory umbrella (Figure 4a, Supplementary Tables 7 & 8).



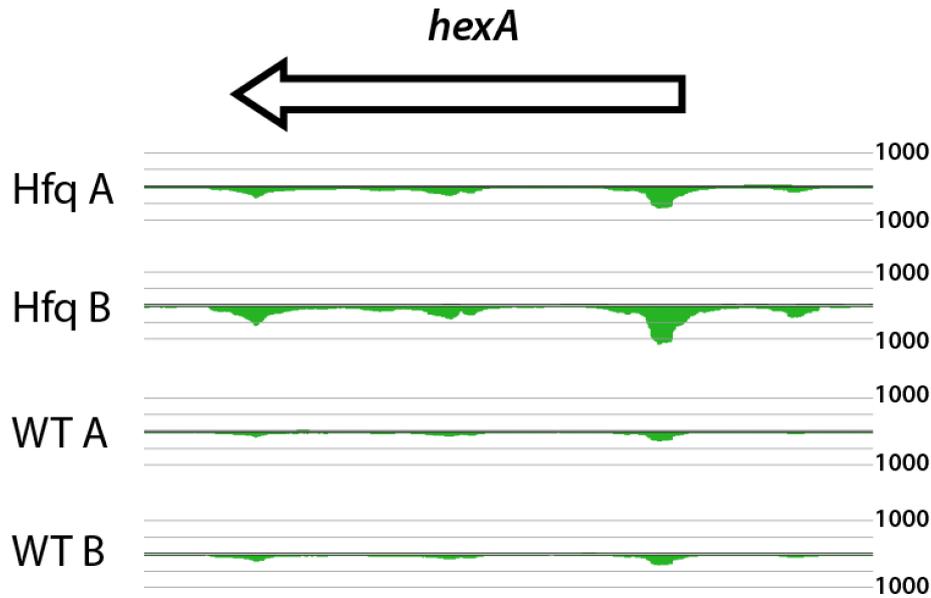
### Supplementary Figure 1

Sequence of region in *P. laumondii* TTO1 containing predicted *arcZ* sequence (green arrow). The 3' end of *arcB* (blue arrow) is also shown. Dotted red lines indicate region of *arcZ* that was deleted. Also indicated is the site of insertion from transposon sequencing (inverted black triangle), as well as the -35 and -10 promoter regions and the transcriptional start site (+1).



### Supplementary Figure 2

**a.** Genes that were significantly affected in the  $\Delta arcZ$  strain and not the  $\Delta hfq$  strain or **b.** affected in both  $\Delta arcZ$  and  $\Delta hfq$  strains. The coding sequences associated with  $\Delta arcZ$  of *P. laumondii* (green),  $\Delta hfq$  (red) or  $\Delta arcZ::arcZ$  (blue) compared to the WT were grouped into eight different categories: specialized metabolites (SM), regulators, virulence, phage related, cell wall, cell processes, hypothetical and unknown based on their annotations. Vertical lines represent the median for each group. Complete lists of regulated genes for *P. laumondii* mutants can be seen in Supplementary Tables 7-8.



### Supplementary Figure 3

RIP-seq enrichment around the region of *hexA* in Hfq::3xFLAG samples (Hfq A & B) and untagged samples (WT A & B). Plots indicate the strand reads map to (bottom = reverse, top = forward). Scale represents perfectly mapped reads. For all enriched regions, see Supplementary Tables 4 and 5.

#### a HexA\_UTR

ATGTTAATTTAATTTGATAGTGCTTACGTAAA**AACACCAGG**TTAGTTAGTAATTA

AAATCAAAAAAAAAAGTGATGAATAAACAATG



#### b HexA\_Pacl\_UTR

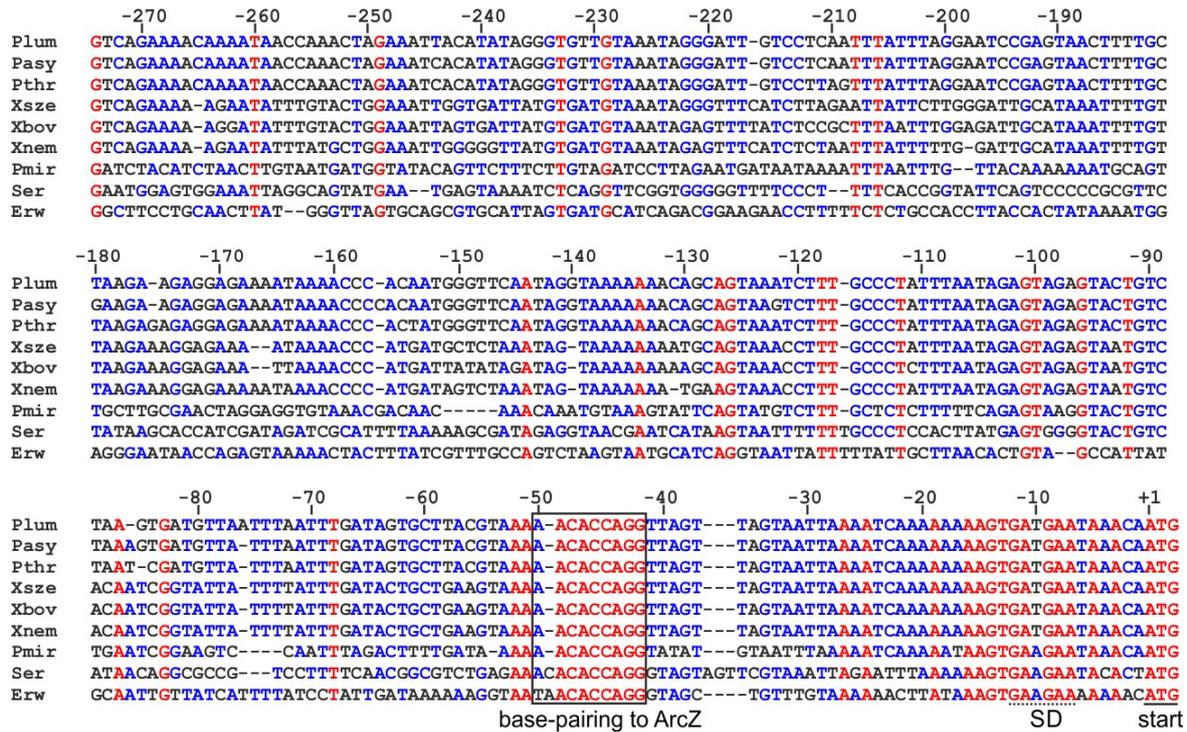
ATGTTAATTTAATTTGATAGTGCTTACGTAAA**TTAATTAA**TTAGTTAGTAATTA

ATCAAAAAAAAAAGTGATGAATAAACAATG



### Supplementary Figure 4

**a.** 5' UTR of *hexA* including the predicted ArcZ binding site (red). The arrow indicates the start of the *hexA* coding sequence. **b.** The predicted ArcZ binding site (AACACCAGG) was exchanged to a *PacI* restriction site (TTAATTAA) as shown.



### Supplementary Figure 5

Alignment of the *hexA* 5' UTR from *P. laumondii* TT01, *P. asymbiotica*, *P. thracensis*, *X. szentirmaii*, *X. bovienii*, *X. nematophila*, *Proteus mirabilis*, *Serratia marcescens* and *Erwinia* sp. J780, beginning with the transcriptional start site. The sequences were aligned using the Multalin Algorithm (21). Black box indicates the region of base-pairing to ArcZ. SD sequence and start codon of *hexA* are underlined. Numbers indicate distance to the start codon.

## References

1. Ettwiller L, Buswell J, Yigit E, & Schildkraut I (2016) A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* 17:199.
2. Tobias NJ, Linck A, & Bode HB (2018) Natural Product Diversification Mediated by Alternative Transcriptional Starting. *Angew Chem Int Ed Engl* 57(20):5699-5702.
3. Tobias NJ, *et al.* (2017) *Photorhabdus*-nematode symbiosis is dependent on hfq-mediated regulation of secondary metabolites. *Environ Microbiol* 19(1):119-129.
4. Yu SH, Vogel J, & Forstner KU (2018) ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. *Gigascience* 7(9).
5. Li L, *et al.* (2013) BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Res* 41:D233-238.
6. Goodman AL, *et al.* (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6(3):279-289.
7. Nollmann FI, *et al.* (2015) A *Photorhabdus* natural product inhibits insect juvenile hormone epoxide hydrolase. *Chembiochem* 16(5):766-771.
8. Lorenzen W, Ahrendt T, Bozhuyuk KA, & Bode HB (2014) A multifunctional enzyme is involved in bacterial ether lipid biosynthesis. *Nat Chem Biol* 10(6):425-427.
9. Waters CM & Bassler BL (2006) The *Vibrio harveyi* quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers. *Genes Dev* 20(19):2754-2767.
10. Thoma S & Schobert M (2009) An improved *Escherichia coli* donor strain for diparental mating. *FEMS Microbiol Lett* 294(2):127-132.
11. Fischer-Le Saux M, Viallard V, Brunel B, Normand P, & Boemare NE (1999) Polyphasic classification of the genus *Photorhabdus* and proposal of new taxa: *P. luminescens* subsp. *luminescens* subsp. nov., *P. luminescens* subsp. *akhurstii* subsp. nov., *P. luminescens* subsp. *laumondii* subsp. nov., *P. temperata* sp. nov., *P. temperata* subsp. *temperata* subsp. nov. and *P. asymbiotica* sp. nov. *Int J Syst Bacteriol* 49 Pt 4:1645-1656.
12. Lengyel K, *et al.* (2005) Description of four novel species of *Xenorhabdus*, family *Enterobacteriaceae*: *Xenorhabdus budapestensis* sp. nov., *Xenorhabdus ehlersii* sp. nov., *Xenorhabdus innexi* sp. nov., and *Xenorhabdus szentirmaii* sp. nov. *Syst Appl Microbiol* 28(2):115-122.
13. Bode E, *et al.* (2019) Promoter Activation in Delta<sub>hfq</sub> Mutants as an Efficient Tool for Specialized Metabolite Production Enabling Direct Bioactivity Testing. *Angew Chem Int Ed Engl* 58(52):18957-18963.
14. Joyce SA, *et al.* (2008) Bacterial biosynthesis of a multipotent stilbene. *Angew Chem Int Ed Engl* 47(10):1942-1945.
15. Brachmann AO, *et al.* (2007) A type II polyketide synthase is responsible for anthraquinone biosynthesis in *Photorhabdus luminescens*. *Chembiochem* 8(14):1721-1728.
16. Bode HB, *et al.* (2012) Determination of the absolute configuration of peptide natural products by using stable isotope labeling and mass spectrometry. *Chemistry* 18(8):2342-2348.
17. Brachmann AO, *et al.* (2013) Pyrones as bacterial signaling molecules. *Nat Chem Biol* 9(9):573-578.
18. Brachmann AO, Forst S, Furgani GM, Fodor A, & Bode HB (2006) Xenofuranones A and B: phenylpyruvate dimers from *Xenorhabdus szentirmaii*. *J Nat Prod* 69(12):1830-1832.
19. Zhou Q, *et al.* (2013) Structure and biosynthesis of xenoamicins from entomopathogenic *Xenorhabdus*. *Chemistry* 19(49):16772-16779.

20. Cai X, *et al.* (2017) Entomopathogenic bacteria use multiple mechanisms for bioactive peptide library design. *Nat Chem* 9(4):379-386.
21. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16(22):10881-10890.

**Additional Supplementary File:**

SI\_Tables\_1-18.xls

Contains Supplementary Tables 1 to 18; available online after publication or upon request.

**Supplementary Table 1.** Significantly affected genes in *X. szentirmaii* ( $\Delta hfq$  RNASeq)

**Supplementary Table 2.** Transcriptional start sites identified for *X. szentirmaii*

**Supplementary Table 3.** OD<sub>600</sub> values of cultures from all experiments

**Supplementary Table 4.** Enriched sRNA sequences from RIP-seq experiment

**Supplementary Table 5.** Enriched mRNA and 5' UTR sequences from RIP-seq experiment

**Supplementary Table 6.** OD<sub>600</sub> of *P. laumondii* WT and TN-mutants after 72 h of cultivation.

**Supplementary Table 7.** Significantly affected genes in TT01  $\Delta hfq$  and  $\Delta arcZ$

**Supplementary Table 8.** Significantly affected genes in TT01 only  $\Delta arcZ$

**Supplementary Table 9.** ArcZ sequences used in Copra analysis

**Supplementary Table 10.** Top 200 hits defined by Copra as potentially interacting with ArcZ

**Supplementary Table 11.** Proteomic analysis of *P. laumondii* WT,  $\Delta arcZ$ ,  $\Delta hfq$  and  $\Delta hexA::hexA\_Pacl\_UTR$

**Supplementary Table 12.** Significantly affected genes in *X. szentirmaii* ( $\Delta arcZ$  RNASeq)

**Supplementary Table 13.** Plasmids used in this study

**Supplementary Table 14.** Strains used in this study

**Supplementary Table 15.** Primer sequences used in this study

**Supplementary Table 16.** Compounds targets used in HPLC-MS analyses

**Supplementary Table 17.** Updated *P. laumondii* TTO1 annotation formatted as .gff

**Supplementary Table 18.** Updated *X. szentirmaii* DSM16338 annotation formatted as .gff

## 5 RIL-seq reveals a quorum sensing small RNA sponge

**Michaela Huber**, Mona Hoyos, Sahar Melamed, Alexandre Vicente, Gisela Storz,  
and Kai Papenfort. “RIL-seq reveals a quorum sensing small RNA sponge.”

Manuscript in preparation.



## **RIL-seq reveals a quorum sensing small RNA sponge**

Michaela Huber<sup>1,2</sup>, Mona Hoyos<sup>1,2</sup>, Sahar Melamed<sup>4</sup>, Alexandre Vicente<sup>2</sup>, Gisela Storz<sup>4</sup>,  
and Kai Papenfort<sup>1,3</sup>

<sup>1</sup> Friedrich Schiller University Jena, Institute of Microbiology, 07745 Jena, Germany.

<sup>2</sup> Faculty of Biology I, Ludwig-Maximilians-University of Munich, 82152 Martinsried, Germany.

<sup>3</sup> Microverse Cluster, Friedrich Schiller University Jena, 07743 Jena, Germany.

<sup>4</sup> Division of Molecular and Cellular Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD 20892-5430, USA.

**ABSTRACT**

Quorum sensing (QS) enables bacteria to coordinate collective behavior in response to cell population density. In *Vibrio cholerae*, QS controls various processes including biofilm formation and virulence. Four homologous small RNAs, Qrr1-4, are central elements of the *V. cholerae* QS system. The Qrr1-4 sRNAs depend on the global RNA chaperone Hfq to engage base-pairing with multiple mRNA targets and carry out their function as post-transcriptional regulators. Hfq plays a key role for stabilizing sRNAs and facilitating their interaction with target mRNAs in many Gram-negative bacteria, including *V. cholerae*. In this study, we employed RIL-seq analysis to decipher the network of sRNA-target pairs mediated by Hfq in this pathogen. We detected thousands of putative interactions, and by confirming various interactions via an established reporter system substantially increased the number of currently known sRNA targets in *V. cholerae*. Furthermore, detailed investigation showed that a novel sRNA, named QrrS, inhibits the functions of the Qrr1-4 sRNAs by destabilizing them upon base-pairing. Our results demonstrate that QrrS acts as a sponge RNA in the QS system when cells transition from low to high cell density. Altogether, our work provides global insights into and the basis for further studies on Hfq-mediated RNA-RNA interactions in *V. cholerae* and highlights the importance of a previously overlooked sRNA sponge in the QS system of this major human pathogen.

**INTRODUCTION**

Quorum sensing (QS) is a very common principle among microorganisms allowing regulation of gene expression in response to changes in cell density (1). QS is a cell-to-cell communication process which is based on the release of and response to extracellular signaling molecules called autoinducers (2). It allows bacteria to switch between individual and collective lifestyles. Various processes, including biofilm formation, virulence and bioluminescence are controlled by QS in diverse species (3, 4).

*Vibrio* species are model organisms for studying QS-controlled behavior (5). In *Vibrio cholerae*, the causative agent of the cholera disease, QS is intimately linked to pathogenesis (6). *V. cholerae* cells produce and respond to three different autoinducer molecules, CAI-1 (cholerae autoinducer-1), AI-2 (autoinducer-2) and DPO (6-9). CAI-1 and AI-2 are detected by the membrane bound receptors CqsS and LuxPQ, respectively, which act as phosphatases upon binding of the autoinducer molecules (6). At low cell density, when the concentration of autoinducer molecules is low, the receptors act as kinases, which leads to transcriptional activation of the four homologous small RNAs (sRNAs) Qrr1-4 via a signal transduction cascade (10). Qrr1-4 activate translation of AphA, the master transcriptional regulator at low cell density, triggering expression of genes for biofilm formation and

virulence (11, 12). Additionally, Qrr1-4 repress *hapR*, encoding the master high cell density regulator (12, 13). Thereby, Qrr1-4 are key elements for orchestrating QS behavior in *V. cholerae*. The autoinducer molecule DPO functions via an independent QS pathway; binding of DPO to the transcription factor VqmA results in activation of the VqmR sRNA (14). VqmR inhibits virulence and biofilm formation at high cell density by repressing multiple targets including *aphA* and *vpsT*, encoding a major transcriptional activator of genes required for biofilm production (14, 15).

Qrr1-4 as well as VqmR are Hfq-dependent sRNAs (10, 13, 14). Hfq is known as a central RNA chaperone in Gram-negative bacteria, promoting base-pairing interactions between *trans*-acting sRNAs and target mRNAs (16, 17). It has recently been shown that Hfq associates with hundreds of transcripts including a large number of sRNA candidates in *V. cholerae* (18), similar to what has previously been reported for the model organisms *Escherichia coli* and *Salmonella enterica* (19-22). The global role of Hfq in post-transcriptional gene regulation is reflected by drastic phenotypes of *hfq* mutants, such as decreased stress tolerance and loss of virulence in many pathogens, including *V. cholerae* (23, 24).

Although hundreds of bacterial sRNAs have been identified and predicted in the last two decades, identification of their targets remains a major challenge (25, 26). Recently, a novel method aiming to identify transcriptome-wide RNA-RNA interactions in bacterial cells mediated by Hfq has been developed. RIL-seq (RNA interaction by ligation and sequencing) allows to capture and decipher the network of RNA duplexes formed on Hfq on a large scale (27, 28). In addition to identifying canonical sRNA-mRNA pairs, the method also facilitates the discovery of interactions between two non-coding RNAs. In the past years, several “sponge” sRNAs have been found which act to titrate other sRNAs away from their mRNA targets, a scenario which might play a previously underestimated role in post-transcriptional gene regulation (29).

In this study, we applied RIL-seq to *V. cholerae* Hfq to get insights into the RNA-RNA “interactome” of this model pathogen. Our analysis revealed approximately 3,000 interactions at two different growth conditions, recovering several previously reported targets of Hfq-binding sRNAs in *V. cholerae*. We discovered a novel sRNA, which base-pairs with the quorum regulatory RNAs Qrr1-4 and which we named QrrS (**Q**rr1-4 **s**ponge). Our results show that QrrS, activated by a LysR-type transcription factor, functions as a sponge RNA by destabilizing the Qrr1-4 sRNAs. Thereby, QrrS acts as an important player in the *V. cholerae* QS system, accelerating transition from low to high cell density mode.

## METHODS

### Bacterial strains and growth conditions

All strains used in this study are listed in Supplementary Table S3. *V. cholerae* and *E. coli* strains were cultivated under aerobic conditions in LB or M9 minimal medium (0.4% glucose, 0.4% casamino acids) at 37°C, unless stated otherwise. Where appropriate, antibiotics were used at the following concentrations: 100 µg/ml ampicillin, 20 µg/ml chloramphenicol, 50 µg/ml kanamycin, 50 U/ml polymyxin B, 5,000 µg/ml streptomycin, and 5 µg/ml tetracycline. For transcript stability experiments, rifampicin was used at 250 µg/ml.

### RIL-seq experiment

*V. cholerae* wild-type and *hfq::3XFLAG* strains were cultivated in duplicates in LB medium to low (OD<sub>600</sub> of 0.2) and high cell densities (OD<sub>600</sub> of 2.0). The experimental as well as the computational part of the RIL-seq method were carried out as described in *Melamed et al.* (27). Briefly, cells corresponding to 50 OD<sub>600</sub> units were subjected to protein-RNA cross-linking, cell lysis and co-immunoprecipitation using anti-FLAG-antibody (Sigma; F1804). Subsequently, the co-immunoprecipitated RNA was treated with RNase A/T1 and T4 RNA ligase. Samples were subjected to proteinase K digestion, and RNA was extracted. Afterwards, RNA was fragmented and treated with TURBO DNase. Ribosomal RNA was depleted and cDNA libraries were prepared. cDNA libraries were sequenced in paired-end mode on a HiSeq 2500 system (Illumina), and reads were mapped to the *V. cholerae* reference genome (NCBI accession numbers NC\_002505.1 and NC\_002506.1), including annotations for Vcr001-Vcr107 (14) and Vcr200-Vcr230 (18). Data analysis was performed according to the previously published computational pipeline (27).

### Fluorescence measurements

GFP fluorescence assays were performed as described previously (30) with *E. coli* Top 10 cells cultivated overnight in LB medium. To measure promoter activity, *V. cholerae* strains carrying an mKate2 transcriptional reporter were cultivated in M9 minimal medium and samples were collected at the indicated time points. For all fluorescence measurements, three independent replicates were used for each strain. Cells were resuspended in PBS and relative fluorescence was determined using a Spark 10 M plate reader (Tecan). Control samples not expressing fluorescent proteins were used to subtract background fluorescence.

**Western blot analysis**

Western blot analysis of FLAG-tagged proteins was performed as described previously (9). In brief, samples were separated using SDS-PAGE and subsequently transferred to PVDF membranes. Anti-FLAG antibody (Sigma; F1804) was used for detection. RNAP served as loading control and was detected using anti-RNAP antibody (BioLegend; WP003). Signals were visualized on a Fusion FX EDGE imager (Vilber Lourmat).

**RNA isolation and Northern blot analysis**

Total RNA samples were prepared as described previously (31). RNA samples were separated on 6% polyacrylamide / 7M urea gels and transferred to Hybond-XL membranes (GE Healthcare) by electro-blotting. Membranes were hybridized in Roti-Hybri-Quick buffer (Roth) at 42°C with [<sup>32</sup>P] end-labeled DNA oligonucleotides. Oligonucleotides used for probing are listed in Supplementary Table S4. Membranes were washed in three subsequent steps with SSC (5x, 1x, 0.5x) / 0.1% SDS wash buffer. Signals were visualized on a Amersham Typhoon phosphorimager (GE Healthcare) and quantified with GelQuant software (BiochemLabSolutions).

**Bioluminescence assay**

*V. cholerae* cells harboring the pBB1 vector (6), which carries the *V. harveyi luxCDABE* operon, were grown overnight in SOC broth (32) supplemented with tetracycline, and diluted 1:1000 into fresh medium. Light production was measured at the indicated time-points during growth of the diluted cultures using a Spark 10 M plate reader (Tecan). Three independent replicates were used for each strain.

**Plasmid construction**

All plasmids and all DNA oligonucleotide sequences are listed in Supplementary Tables S2 and S4, respectively. GFP reporter fusions were constructed as previously described (30), and using previously determined transcriptional start sites (14). The pXG10 vector was used for monocistronic genes, the pXG30 vector for operons (30). Inserts were amplified from *V. cholerae* genomic DNA with the respective oligonucleotide combinations indicated in the following and cloned into linearized pXG10 (KPO-1702/KPO-1703) via Gibson assembly (GA) (33): pMH063 (KPO-4210/KPO-4211), pJR026 (KPO-3795/KPO-3796), pJR039 (KPO-4137/KPO-4138), pKT006 (KPO-5411/KPO-5412), pKT001 (KPO-5191/KPO-5192), pMH073 (KPO-4937/KPO-4938), pJR040 (KPO-4132/KPO-4133), pKT005 (KPO-5408/KPO-5367), pMH064 (KPO-4644/KPO-4645), pKT008 (KPO-5409/KPO-5369), pKT007 (KPO-5410/KPO-5371), pMH067 (KPO-4642/KPO-4643), pMH060 (KPO-4276/KPO-4277), pMD092 (KPO-2573/KPO-2574), pMH058 (KPO-4058/KPO-4059),

pJR029 (KPO-3813/KPO-3814), pMH071 (KPO-4056/KPO-4057), pMH062 (KPO-4212/KPO-4213), pJR044 (KPO-4471/KPO-4472), pJR043 (KPO-4469/KPO-4470), pJR042 (KPO-4184/KPO-4185), pMH059 (KPO-4078/KPO-4079), pMH056 (KPO-4060/KPO-4061), pJR045 (KPO-4473/KPO-4474), pMD161 (KPO-2779/KPO-2780), pJR036 (KPO-4019/KPO-4020). For pYH034 (KPO-3005/KPO-3006) and pYH033 (KPO-3003/KPO-3004), pXG10 and respective inserts were digested with NsiI and NheI and ligated. pNP064 (KPO-1720/KPO-1721), pNP058 (KPO-1708/KPO-1709) and pYH038 (KPO-3054/KPO-3055) were constructed likewise, using pXG10-1C (34). For pXG30 fusions, backbone was linearized with KPO-4646/KPO-1703, and inserts, amplified with the indicated oligonucleotide combinations, were fused via GA: pKT002 (KPO-5189/KPO-5190), pKT003 (KPO-5187/KPO-5188), pMH066 (KPO-4651/KPO-4136), pMH072 (KPO-4935/KPO-4936), pMH068 (KPO-4647/KPO-4648), pMH069 (KPO-4929/KPO-4630), pMH065 (KPO-4649/KPO-4650), pKT004 (KPO-5209/KPO-5410). Constitutive sRNA expression plasmids pMD104, pMH057, pSG001 and pJR035 were constructed by PCR amplification of the respective sRNA genes from *V. cholerae* genomic DNA using oligonucleotide combinations KPO-2570/KPO-2565, KPO-4062/KPO-4063, KPO-1858/KPO-1859 and KPO-3965/KPO-3966, respectively, and cloning via GA into pEVS143 (35) vector backbone, linearized with KPO-0092/KPO-1397. For plasmids pRH002, pRH003, pRH006, pNP004 and pAS001, inserts were obtained by PCR amplification using KPO-1092/KPO-1093, KPO-1090/KPO-1091, KPO-1084/KPO-1085, KPO-1005/KPO-1006 and KPO-1076/KPO-1077, respectively, and introduced into linearized pEVS143 plasmid backbone (KPO-0092/KPO-1023) using XbaI restriction and ligation. pMD099 and pMD176 were constructed by amplifying the *qrrS* and the *qrr4* genes from *V. cholerae* genomic DNA using oligonucleotides KPO-2558/KPO-2559 and KPO-3779/KPO-3780, respectively, and cloning via GA into pBAD1K (pMD004), linearized with KPO-0196/KPO-1397. Plasmid pMD103 was obtained by site-directed mutagenesis of pMD099 using KPO-3749 and KPO-3750. For pAF012, the promoter region of QrrS was amplified from *V. cholerae* genomic DNA with KPO-3676/KPO-3677 and GA was used to insert it into the pCMW-1C-mKate vector (15), linearized with KPO-2591/KPO-2592. For pKAS32 (36) plasmids pAF013 and pMH075, backbone was linearized with KPO-0167 and KPO-0168, and inserts were fused using GA. For pAF013, up and down flanks were amplified from genomic DNA using KPO-3741/KPO-2742 and KPO-3743/KPO-2743; for pMH075, the insert (*vca0830* coding sequence and 3XFLAG sequence) was generated via IDT gene block synthesis. Plasmid pAS005 was constructed by amplifying the up and down flanks of the *qrrS* gene from genomic DNA with KPO-1301/KPO-1304 and KPO-1302/KPO-1305, and cloning them into pKAS32 via restriction digest with KpnI and AvrII and ligation.

### Strain construction

All strains used in this study are listed in Supplementary Table S3. *V. cholerae* C6706 was used as wild-type strain. RK2/RP4-based conjugal transfer served to introduce plasmids from *E. coli* S17 $\lambda$ pir donor strains into *V. cholerae*. Subsequently, transconjugants were selected using appropriate antibiotics and polymyxin B to specifically inhibit *E. coli* growth. All *V. cholerae* mutant strains were constructed using the pKAS32 suicide vector (36). Briefly, pKAS32-plasmids were conjugated into *V. cholerae* and cells were selected for ampicillin resistance. Single colonies were streaked on fresh plates and selected for streptomycin resistance. Desired mutations were confirmed by PCR and sequencing.

## RESULTS

### RIL-seq analysis of Hfq in *V. cholerae*

To globally identify RNA-RNA interactions mediated by Hfq in *V. cholerae*, we applied the RIL-seq protocol (27) to *V. cholerae* wild-type cells carrying a chromosomal 3XFLAG epitope at the C-terminus of the *hfq* gene (*vc0347*). Hfq has previously been shown to be expressed at low as well as at high cell densities in *V. cholerae* (18). To capture RNA-RNA interactions at different stages of growth, we performed the RIL-seq experiments with cells cultivated to exponential (OD<sub>600</sub> of 0.2) and to stationary phase (OD<sub>600</sub> of 2.0). *V. cholerae* wild-type cells lacking the 3XFLAG epitope were used as negative control. Reads from deep sequencing of the cDNA libraries were mapped to the *V. cholerae* N16961 genome (37). We obtained a total of ~130 million reads for the IP samples and ~153 million reads for the total RNA samples (corresponding to two replicates of each strain and condition) with on average 94% and 92% mapping, respectively. We combined the libraries from the two replicates to unified data sets resulting in approximately 3,000 significant chimeric interactions at both growth conditions in the Hfq::3XFLAG strain (Fig. 1 A and B). In contrast, the number of significant chimeras in the untagged control was negligible (7 at OD<sub>600</sub> of 0.2, 14 at OD<sub>600</sub> of 2.0), indicating that the chimeric fragments were specifically bound to Hfq. Several previously validated interactions in *V. cholerae* were recovered in our experiments (Table S1), supporting our results.

Next, we focused on nine Hfq-binding sRNAs (Qrr2, Qrr4, Spot 42, GcvB, TfoR, Vcr001, Vcr017, Vcr043, Vcr227) and tested 56 predicted mRNA targets. To do so, we used a well-established reporter system, where the sRNA of interest is expressed constitutively from one plasmid and a translational fusion of the putative target to *gfp* is expressed from a second plasmid (30). We co-transformed the respective plasmids in *E. coli* Top 10 cells and tested target regulation by measuring GFP levels. With this approach, we confirmed post-

transcriptional repression (fold-change  $\geq 2$ ) of 34 targets (Fig. 1C). Two targets (*rbmC* and *zapA*) showed upregulation upon overexpression of the respective sRNAs (Fig. 1C). These results significantly amplify the number of validated sRNA-mRNA interactions in *V. cholerae*.

### **QrrS base-pairs with and destabilizes the quorum sensing regulatory RNAs Qrr1-4**

In addition to many sRNA-mRNA interactions, our data uncovered several putative sRNA-sRNA interactions. Most interestingly, we detected that one sRNA, previously identified as Vcr103 (14), was involved in chimeric fragments with all four quorum sensing regulatory RNAs (Qrr1-4), but no other transcripts (Fig. 2A). Thus, we hypothesized that the sRNA could act as a sponge of the Qrr sRNAs and therefore termed it QrrS (**Qrr1-4 sponge**). Northern blot analysis indicated that QrrS is expressed at all stages of growth, being most abundant at stationary phase (Fig. 2B). To test the hypothesis that QrrS could base-pair with the Qrr1-4 sRNAs, we used the *RNA-hybrid* algorithm (38) to predict base-pairing interactions, and indeed found extensive regions of complementarity between QrrS and Qrr1-4 (Fig. 2C). Alignment of *qrrS* sequences from various *Vibrio* species showed that the region of predicted base-pairing is highly conserved (Fig. 2D), supporting its involvement in a regulatory function.

Subsequently, we aimed to explore the regulatory effect of QrrS on the Qrr1-4 sRNAs. To this end, we constructed a plasmid where expression of QrrS can be induced with L-arabinose from the P<sub>BAD</sub> promoter. We cultivated *V. cholerae* wild-type cells harboring either pBAD-*qrrS* or a control plasmid (pBAD-ctr) to exponential phase (OD<sub>600</sub> of 0.2), induced expression of QrrS for 15 minutes and then added rifampicin to stop transcription initiation. Northern blot analysis served to monitor stability of Qrr4, used as a proxy for the four homologous Qrr sRNAs. Whereas in the control, half-life of Qrr4 was  $> 32$  min (Fig. 3A, lanes 1-6), it was reduced to  $< 2$  min upon overexpression of QrrS (Fig. 3A, lanes 7-12), indicating that QrrS acts to destabilize the Qrr sRNAs. To test if this effect is mediated by the predicted region of base-pairing, we inserted a single point mutation in the *qrrS* gene (G72C) (Fig. 2C). Using the mutated plasmid (pBAD-*qrrS*\*), the effect on Qrr4 was abolished (Fig. 3A, lanes 13-18). Hence, we conclude that QrrS destabilizes the Qrr1-4 sRNAs upon base-pairing via the predicted region. To investigate if the Qrr1-4 sRNAs have a similar effect on QrrS, we generated a plasmid where Qrr4 can be induced from the inducible P<sub>BAD</sub> promoter and repeated the experiment vice versa. Here, we observed only a very modest reduction of QrrS stability upon Qrr4 overexpression (Fig. 3B). However, processing products became more prominent (Fig. 3B). Overall, the results suggest a directionality of regulation where QrrS strongly destabilizes the Qrr1-4 sRNAs while QrrS itself is not destabilized to the same extent.

### **QrrS is activated by the LysR-type transcription factor VCA0830**

The *qrrS* gene is located antisense downstream of the *vca0830* gene encoding a LysR-type transcription factor (Fig. 4A). We therefore speculated that this transcription factor might play a role for the expression of QrrS. To test this, we constructed a *V. cholerae* mutant strain with a deletion of the *vca0830* gene and monitored expression of QrrS on Northern blot. Indeed, QrrS levels were clearly reduced (~threefold) at all stages of growth (Fig. 4B), suggesting that VCA0830 activates transcription of QrrS. Moreover, we constructed a mKate2-based transcriptional reporter for QrrS and determined production of the fluorescent protein in wild-type and  $\Delta vca0830$  cells. Consistent with the expression of QrrS levels observed by Northern blot analysis, QrrS promoter activity was strongly increased at stationary phase and was significantly reduced in  $\Delta vca0830$  cells at all tested growth conditions (Fig. 4C).

To examine expression of the VCA0830 protein over growth, we added a chromosomal 3XFLAG epitope to the C-terminus of the *vca0830* gene and analyzed protein levels on Western blot. In contrast to QrrS levels (Fig. 4B), VCA0830 protein levels did not increase significantly at stationary phase (Fig. 4D). For LysR-type transcriptional regulators it has been reported that binding of a small molecule to the co-inducer binding domain can induce a structural change altering the DNA binding capacities (39). We hence speculate that there might be an additional factor involved binding to VCA0830 and thereby controlling transcriptional regulation of QrrS.

### **QrrS acts as a sponge in the quorum sensing system of *V. cholerae***

Although we have not yet understood in detail how expression of QrrS is regulated, our data suggest that QrrS acts as a sponge in the quorum sensing system of *V. cholerae*. To explore this role in more detail, we constructed a *V. cholerae* mutant with a deletion of the *qrrS* gene. In order to test the effect of the deletion on the quorum sensing behavior, we used a heterologous system based on the *V. harveyi lux* (luciferase) operon. In *V. harveyi*, the operon is activated by the transcription factor LuxR and drives bioluminescence in response to QS signals (40). When expressed ectopically in *V. cholerae*, the operon can be activated by the major QS transcriptional regulator HapR, and thereby allows to examine QS behavior by using light production as read-out (6). We cultivated *V. cholerae* wild-type,  $\Delta qrr1-4$  and  $\Delta qrrS$  strains and monitored light production at different cell densities. In line with previous results (6, 10), wild-type cells were strongly bioluminescent at high cell density and showed low levels of light production at low cell density, whereas  $\Delta qrr1-4$  cells, locked in high cell density state, were constantly bright (Fig. 5). We observed that  $\Delta qrrS$  cells behaved similar to wild-type cells when switching from high to low cell density, but were strongly delayed in the reverse direction, in the transition from low to high cell density (Fig. 5). This suggests

that QrrS acts as a sponge in the QS system of *V. cholerae*, accelerating the transition from low to high cell density.

## DISCUSSION

Identification of targets of bacterial sRNAs is a major challenge to understand post-transcriptional gene regulation. RIL-seq and similar techniques such as CLASH (41, 42) offer new possibilities to study sRNA-target interactions globally. Here, RIL-seq has been applied to *V. cholerae* Hfq, thereby providing the first comprehensive study of the Hfq mediated RNA-RNA network in this model organism and a valuable basis for further studies on Hfq-dependent sRNAs in this organism.

Recent research has shown that sRNA-based regulation is not restricted to the canonical model of activating or repressing mRNA targets. Instead, base-pairing interactions which do not necessarily result in up- or downregulation of expression levels and potentially fulfill other functions seem to be more widespread than previously assumed (42, 43). Hence, an important future challenge will be to identify these true interactions and distinguish them from false positives in such high-throughput data sets. In this context, it might also be interesting to further analyze binding motifs and base-pairing predictions on a large scale.

In addition to identifying sRNA-mRNA interactions, RIL-seq and CLASH have repeatedly detected previously unknown sRNA-sRNA interactions. For instance, RIL-seq applied to *E. coli* Hfq identified a putative 3' UTR derived sRNA, named PspH, as a sponge of the Spf sRNA (27). CLASH applied to *E. coli* Hfq revealed that ArcZ acts as a sponge of the CyaR sRNA (42). Although the exact biological functions here often remain unclear, this might indicate that base-pairing between two non-coding RNAs might be a previously underestimated factor for RNA-based gene regulation networks.

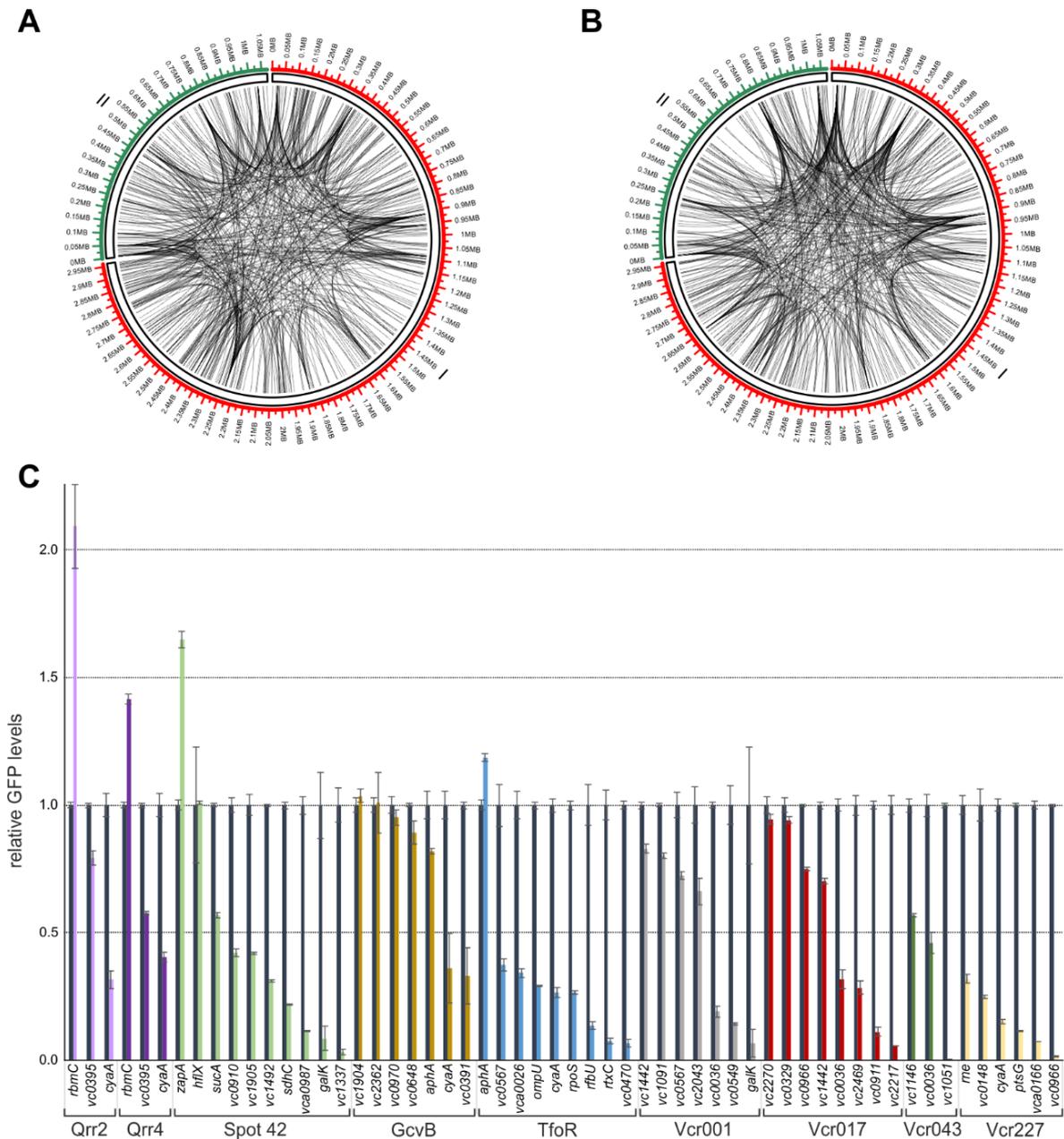
Here, we identified a hitherto overlooked sponge sRNA in the quorum sensing system of *V. cholerae*. Our data indicate that QrrS base-pairs with the four homologous Qrr1-4 sRNAs and that the outcome of the interaction is directional, leading to rapid degradation of Qrr1-4 but not of QrrS (Fig 3). This is similar to what has been observed for the ArcZ-CyaR pair, resulting in an unidirectional regulation of CyaR levels (42). For QrrS-Qrr4, it would be interesting to further investigate the molecular mechanism and potential RNases involved in degradation. Primer extension experiments could help here to determine possible cleavage sites. Interestingly, the site of base-pairing in Qrr1-4 with QrrS overlaps with the site of base-pairing of Qrr1-4 with the *hapR* mRNA, one of the well-studied mRNA targets of Qrr1-4 in *V. cholerae* (13). In this interaction, however, binding of Qrr1-4 induces destabilization of the mRNA target (13, 44). Hence, an exciting future question to address is what exactly

determines the degradation upon pairing with the respective partner RNA in one case but not in the other.

Another aspect which will require further investigation is the exact transcriptional control of QrrS. Our data indicate that expression of QrrS is dependent on the LysR-type transcription factor encoded by the *vca0830* gene (Fig. 4A). However, we currently do not know which conditions exactly determine binding to the promoter of QrrS. As expression of the protein does not show significant differences over growth (Fig. 4D), we speculate that an intra- or extracellular molecule might bind to VCA0830 triggering its DNA binding affinity. To examine if a secreted molecule is binding to the protein, cell-free supernatants could be tested for their ability to activate transcription. Furthermore, a potential molecule could be co-purified together with the transcription factor. Identification of such a molecule clearly would further enhance our understanding of the role QrrS plays in the *V. cholerae* QS system.

Nevertheless, our results clearly show that QrrS is a central player influencing QS dynamics in *V. cholerae* (Fig. 5). As QrrS acts as a sponge for Qrr1-4 (Fig. 3), which regulate the QS master regulators AphA and HapR (12, 13), we believe that QrrS has a global influence on gene expression in *V. cholerae*. We anticipate elevated AphA and decreased HapR levels in a  $\Delta qrrS$  strain, affecting biofilm formation and expression of virulence genes. Thus, we plan to perform transcriptome studies in the future to get global insights into altered gene expression in  $\Delta qrrS$  cells and fully understand its regulatory impact.

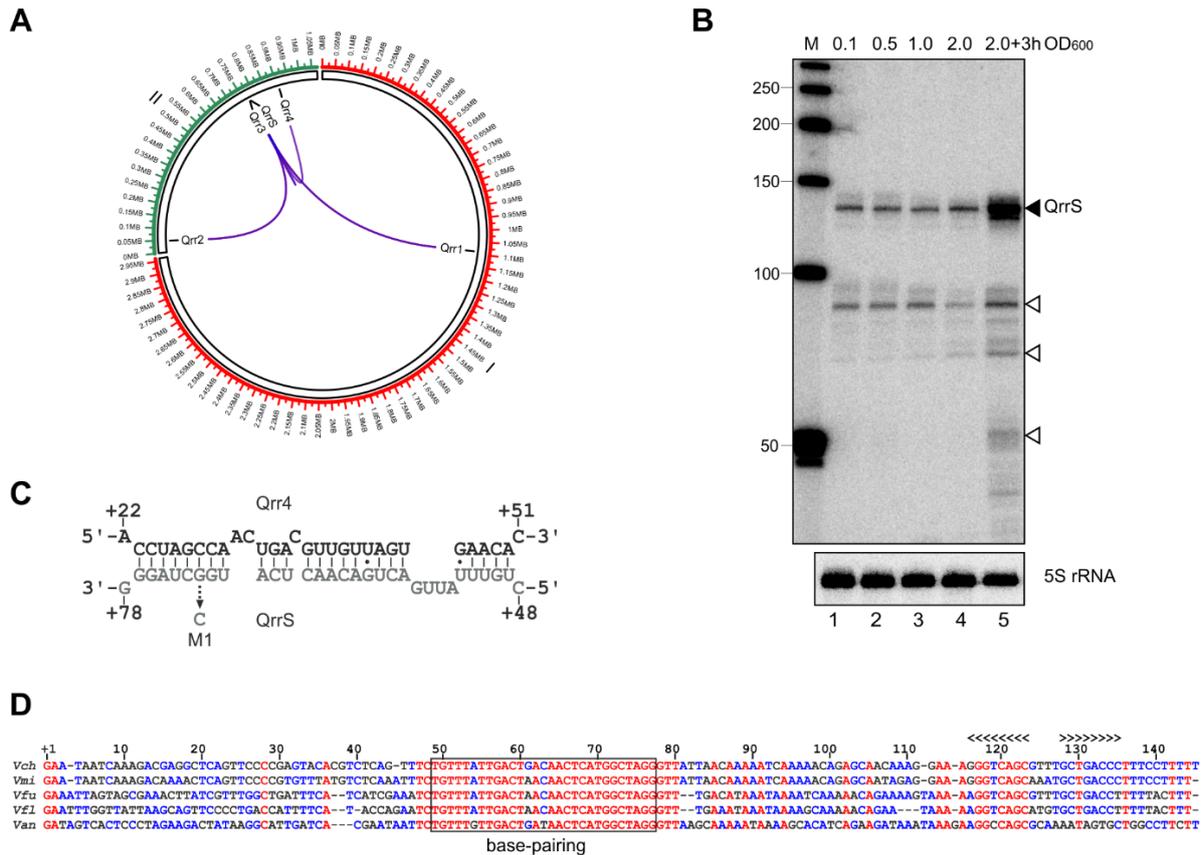
## FIGURES AND LEGENDS



**Figure 1: RIL-seq analysis of Hfq in *V. cholerae***

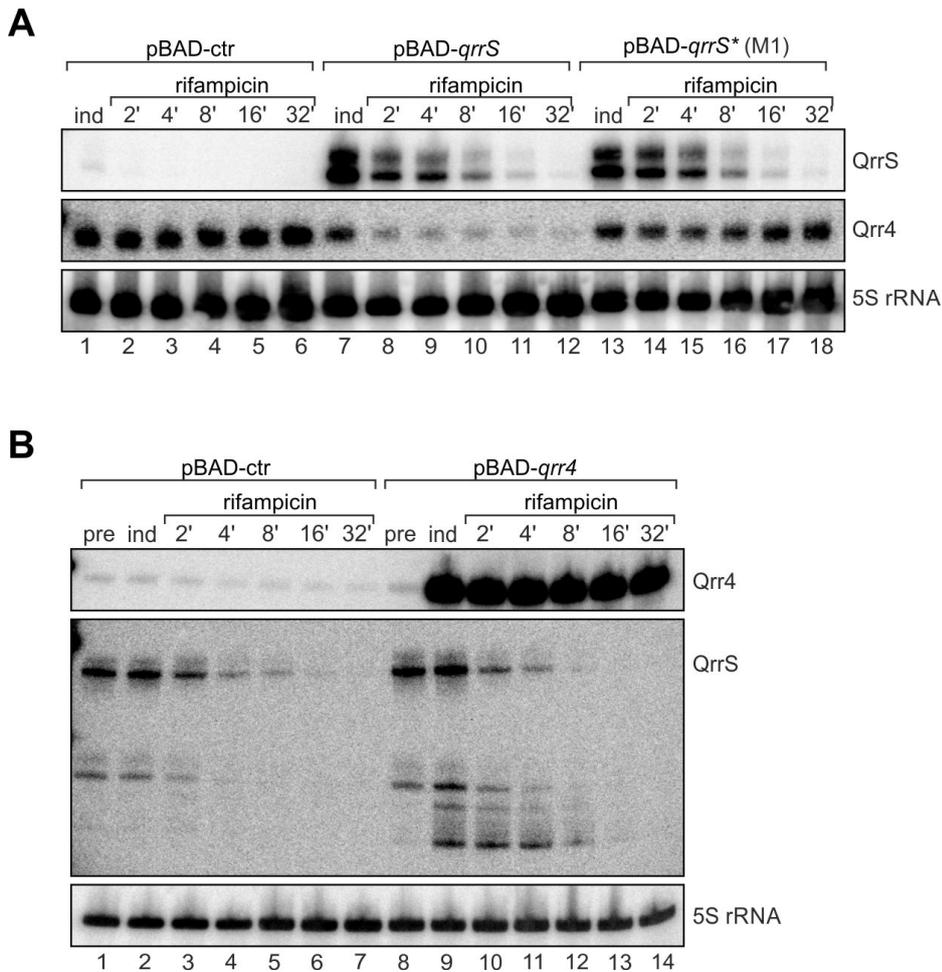
**A** and **B**) Circos plots visualizing Hfq-mediated RNA-RNA interactions. *V. cholerae* *hfq::3XFLAG* cells were cultivated to low (OD<sub>600</sub> of 0.2) (A) and high cell densities (OD<sub>600</sub> of 2.0) (B) and subjected to RIL-seq analysis. Top 500 significant chimeras are shown. The first and the second chromosome are marked in red and green and with I and II, respectively.

**C**) Validation of RNA-RNA interactions predicted by RIL-seq. Translational GFP reporter fusions were co-transformed with a constitutive sRNA expression plasmid or an empty control plasmid in *E. coli* Top 10 cells. GFP production was measured and fluorophore levels from the control strains were set to 1. Error bars indicate SD of three biological replicates.



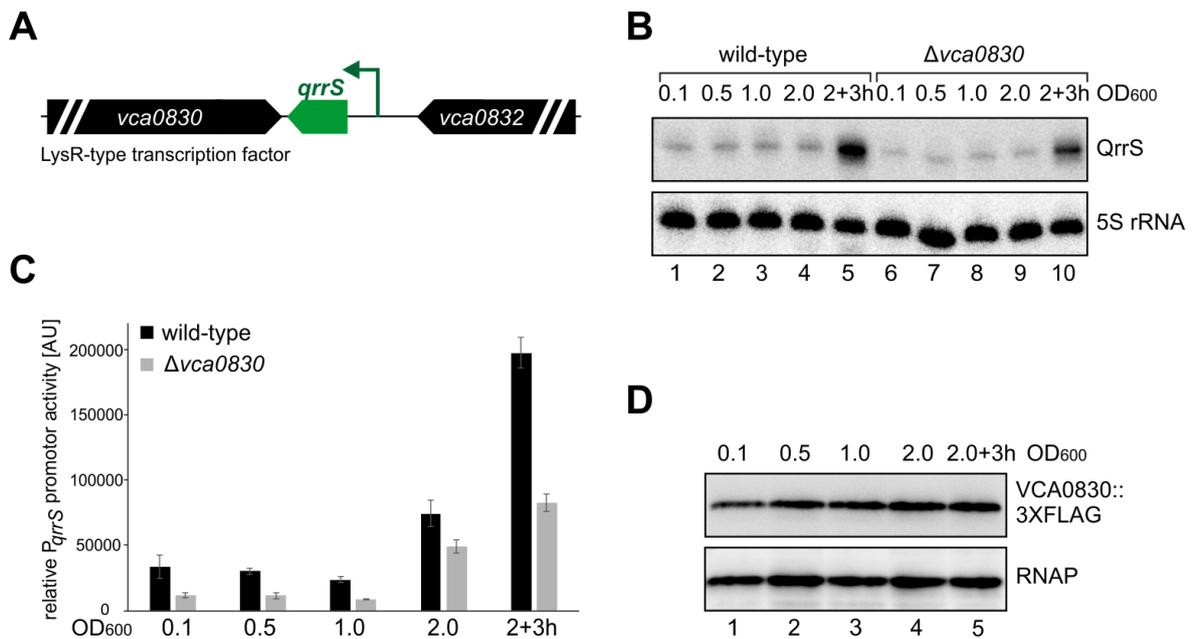
**Figure 2: QrrS base-pairs with the Qrr1-4 sRNAs**

**A)** Circos plot visualizing interaction partners of QrrS. All significant chimeras (cutoff > 13 chimeras) are shown. The first and the second chromosome are marked in red and green and with I and II, respectively. **B)** Northern blot analysis of QrrS expression. *V. cholerae* wild-type cells were cultivated in M9 minimal medium and total RNA samples were collected at the indicated time points. Closed triangles indicate full length QrrS, open triangles indicate processing products. 5S rRNA served as loading control. **C)** Predicted base-pairing interaction of QrrS with Qrr4, used as a proxy for the four homologous Qrr sRNAs. Arrow indicates the point mutation tested in Fig. 3A. **D)** Alignment of *qrrS* sequences from various *Vibrio* species. Sequences were aligned using the Multalin tool (45). The Rho-independent terminator and the predicted base-pairing interaction with the Qrr1-4 sRNAs are indicated. *Vch*, *Vibrio cholerae*; *Vmi*, *Vibrio mimicus*; *Vfu*, *Vibrio furnissii*; *Vfl*, *Vibrio fluvialis*; *Van*, *Vibrio anguillarum*.



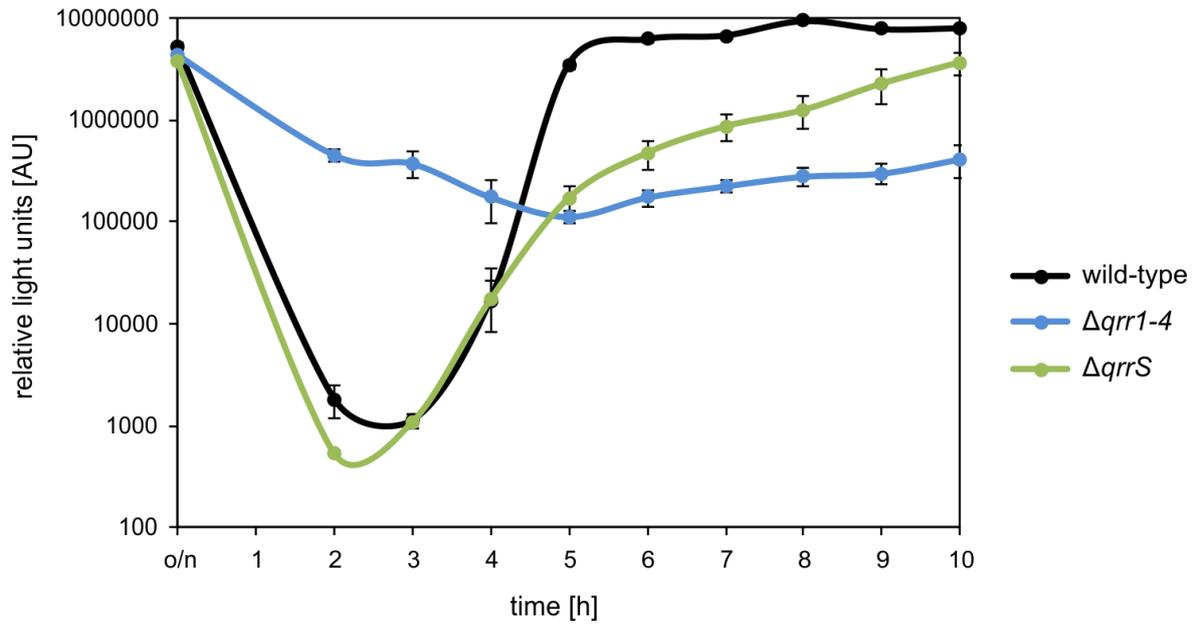
**Figure 3: QrrS destabilizes the Qrr1-4 sRNAs**

**A)** *V. cholerae* cells harboring either a control plasmid (pBAD-ctr), pBAD-*qrrS* or pBAD-*qrrS*\* (carrying a point mutation in the *qrrS* gene as shown in Fig. 2C), were cultivated in LB medium to OD<sub>600</sub> of 0.2. Expression of QrrS was induced with L-arabinose (0.2% final concentration). After 15 minutes of induction, rifampicin was added and total RNA samples were collected at the indicated time points. Northern blot analysis was performed to monitor QrrS and Qrr4 levels. 5S rRNA was used as loading control. **B)** *V. cholerae* cells harboring either a control plasmid (pBAD-ctr) or pBAD-*qrr4* were cultivated in LB medium to OD<sub>600</sub> of 1.0. Expression of Qrr4 was induced with L-arabinose (0.2% final concentration). After 15 minutes of induction, rifampicin was added and total RNA samples were collected at the indicated time points. Northern blot analysis was performed to monitor Qrr4 and QrrS levels. 5S rRNA was used as loading control.



**Figure 4: The transcription factor VCA0830 activates expression of QrrS**

**A)** Schematic representation of the genomic context of *qrrS*. **B)** Northern blot analysis of QrrS expression in *V. cholerae* wild-type and  $\Delta vca0830$  strains. Cells were cultivated in M9 minimal medium and total RNA samples were collected at the indicated time points. 5S rRNA served as loading control. **C)** Measurements of *P<sub>qrrS</sub>* promoter activity. *V. cholerae* wild-type and  $\Delta vca0830$  cells carrying a transcriptional reporter for QrrS (*P<sub>qrrS</sub>::mKate2*) were cultivated in M9 minimal medium. Samples were collected at various stages of growth and analyzed for fluorescence. Error bars represent SD of three independent biological replicates. **D)** Western blot analysis of VCA0830 expression. *V. cholerae* cells carrying a chromosomal 3XFLAG epitope at the *vca0830* gene were cultivated in M9 minimal medium and protein samples were collected at the indicated time points. RNAP served as loading control.



**Figure 5: QrrS alters QS dynamics**

*V. cholerae* wild-type,  $\Delta qrr1-4$  and  $\Delta qrrS$  strains harboring the *V. harveyi luxCDABE* operon on a plasmid (pBB1) were cultivated in SOC broth, supplemented with tetracycline. Light production was measured at the indicated time points. Error bars indicate standard deviation of three independent biological replicates.

## SUPPLEMENTARY INFORMATION

**Table S1:** Previously published sRNA targets recovered in the RIL-seq data set

sRNA	target	repression (-) / activation (+)	reference
VqmR	<i>vca0952 (vpsT)</i>	-	(14)
VqmR	<i>vca0068</i>	-	(14)
VqmR	<i>vc1865</i>	-	(14)
VqmR	<i>vc1063</i>	-	(14)
VqmR	<i>vc1451 (rtxA)</i>	-	(14)
VrrA	<i>vc2213 (ompA)</i>	-	(46)
VrrA	<i>vc1854 (ompT)</i>	-	(47)
VrrA	<i>vca0059 (lpp)</i>	-	(34)
VrrA	<i>vc0429</i>	-	(34)
TfoR	<i>vc1153 (tfoX)</i>	+	(48)
MicX	<i>vc0972</i>	-	(49)
MicX	<i>vc0620</i>	-	(49)
Qrr sRNAs	<i>vc1021 (luxO)</i>	-	(44)
Qrr sRNAs	<i>vc2647 (aphA)</i>	+	(11, 12)
Qrr sRNAs	<i>vc0583 (hapR)</i>	-	(13)
Qrr sRNAs	<i>vca0939</i>	+	(50)
FarS	<i>vc1740 (fadE)</i>	-	(18)
CarZ	<i>vc2390 (carA)</i>	-	(51)
OppZ	<i>vc1092 (oppB)</i>	-	(51)
VadR	<i>vca1075 (crvA)</i>	-	Herzog <i>et al.</i> , in revision
VadR	<i>vc0916 (vpsU)</i>	-	Herzog <i>et al.</i> , in revision
VadR	<i>vc2352</i>	-	Herzog <i>et al.</i> , in revision

**Table S2:** Plasmids used in this study

Plasmid trivial name	Plasmid Stock name	Relevant fragment	Comment	Origin, marker	Reference
<b>Plasmids for RIL-seq target validation (GFP reporter plasmids)</b>					
pXG10-sfGFP	pXG10-sfGFP	lacZ':::sfGFP	Template plasmid for translational reporter	pSC101*, Cm <sup>R</sup>	(30)
pXG30-sfGFP	pXG30-sfGFP	FLAG::lacZ':::sfGFP	Template plasmid for translational reporter (for operons)	pSC101*, Cm <sup>R</sup>	(30)
pXG10-vc0395	pMH063	5' UTR + 20 aa of <i>vc0395</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0122 ( <i>cyaA</i> )	pJR026	5' UTR + 20 aa of <i>vc0122</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc2478 ( <i>zapA</i> )	pJR039	5' UTR + 20 aa of <i>vc2478</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc1905	pKT006	5' UTR + 20 aa of <i>vc1905</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc1492	pKT001	5' UTR + 20 aa of <i>vc1492</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc2091 ( <i>sdhC</i> )	pMH073	5' UTR + 20 aa of <i>vc2091</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0987	pJR040	5' UTR + 20 aa of <i>vc0987</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc1904	pKT005	5' UTR + 20 aa of <i>vc1904</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0970	pMH064	5' UTR + 20 aa of <i>vc0970</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0648	pKT008	5' UTR + 20 aa of <i>vc0648</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc2647 ( <i>aphA</i> )	pKP462	5' UTR + 20 aa of <i>vc2647</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	(15)
pXG10-vc0391	pKT007	5' UTR + 20 aa of <i>vc0391</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0567	pYH034	5' UTR + 15 aa of <i>vc0567</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0534 ( <i>rpoS</i> )	pKP479	5' UTR + 20 aa of <i>vc0534</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0260 ( <i>rfbU</i> )	pYH033	5' UTR + 15 aa of <i>vc0260</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc1449 ( <i>rtxC</i> )	pKP353	5' UTR + 24 aa of <i>vc1449</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	(14)
pXG10-vc0470	pMH067	5' UTR + 20 aa of <i>vc0470</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc1442	pMH060	5' UTR + 20 aa of <i>vc1442</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc1091	pMD092	5' UTR + 20 aa of <i>vc1091</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc2043	pMH058	5' UTR + 20 aa of <i>vc2043</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0036	pJR029	5' UTR + 20 aa of <i>vc0036</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0549	pMH071	5' UTR + 20 aa of <i>vc0549</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0966	pMH062	5' UTR + 20 aa of <i>vc0966</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc2469	pJR044	5' UTR + 20 aa of <i>vc2469</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc0911	pJR043	5' UTR + 20 aa of <i>vc0911</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc2217	pJR042	5' UTR + 20 aa of <i>vc2217</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study

pXG10-vc1146	pMH059	5' UTR + 20 aa of <i>vc1146</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc1051	pMH056	5' UTR + 20 aa of <i>vc1051</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc2030 ( <i>rne</i> )	pJR045	5' UTR + 20 aa of <i>vc2030</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-vc2013 ( <i>ptsG</i> )	pMD161	5' UTR + 20 aa of <i>vc2013</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10- <i>vca0166</i>	pJR036	5' UTR + 20 aa of <i>vca0166</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-1C- <i>vc0930</i> ( <i>rbmC</i> )	pNP064	5' UTR + 15 aa of <i>vc0930</i> ( <i>rbmC</i> )	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-1C- <i>vc0910</i>	pNP058	5' UTR + 15 aa of <i>vc0910</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-1C- <i>vca0026</i>	pYH038	5' UTR + 15 aa of <i>vca0026</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG10-1C- <i>vc0633</i> ( <i>ompU</i> )	pNP085	5' UTR + 15 aa of <i>vc0633</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	(34)
pXG30- <i>vc0348/vc0349</i> ( <i>hflX</i> )	pKT002	3' part of <i>vc0348</i> + IGR + 20 aa of <i>vc0349</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG30- <i>vc2088/vc2087</i> ( <i>sucA</i> )	pKT003	3' part of <i>vc2088</i> + IGR + 20 aa of <i>vc2087</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG30- <i>vc1596/vc1595</i> ( <i>galK</i> )	pMH066	3' part of <i>vc1596</i> + IGR + 20 aa of <i>vc1595</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG30- <i>vc1336/vc1337</i>	pMH072	3' part of <i>vc1336</i> + IGR + 20 aa of <i>vc1337</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG30- <i>vc2363/vc2362</i>	pMH068	3' part of <i>vc2363</i> + IGR + 20 aa of <i>vc2362</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG30- <i>vc2271/vc2270</i>	pMH069	3' UTR of <i>vc2271</i> + IGR + 20 aa of <i>vc2270</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG30- <i>vc0328/vc0329</i>	pMH065	3' part of <i>vc0328</i> + IGR + 20 aa of <i>vc0329</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
pXG30- <i>vc0147/vc0148</i>	pKT004	3' part of <i>vc0147</i> + IGR + 20 aa of <i>vc0148</i>	Translational GFP reporter	pSC101*, Cm <sup>R</sup>	This study
<b>Plasmids for RIL-seq target validation (sRNA expression plasmids)</b>					
p-ctr	pCMW-1		Control plasmid	p15A, Kan <sup>R</sup>	(52)
pEVS143	pEVS143	<i>Ptac</i> promoter	Constitutive overexpression plasmid (template)	p15A, Kan <sup>R</sup>	(35)
p- <i>qrr2</i>	pRH002	<i>qrr2</i>	<i>qrr2</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
p- <i>qrr4</i>	pRH003	<i>qrr4</i>	<i>qrr4</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
p- <i>spot 42</i>	pAS001	<i>spot 42</i>	<i>spot 42</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
p- <i>gcvB</i>	pRH006	<i>gcvB</i>	<i>gcvB</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
p- <i>tfoR</i>	pMD104	<i>tfoR</i>	<i>tfoR</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
p- <i>vcr001</i>	pMH057	<i>vcr001</i>	<i>vcr001</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
p- <i>vcr017</i>	pSG001	<i>vcr017</i>	<i>vcr017</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
p- <i>vcr043</i>	pNP004	<i>vcr043</i>	<i>vcr043</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
p- <i>vcr227</i>	pJR035	<i>vcr227</i>	<i>vcr227</i> expression plasmid	p15A, Kan <sup>R</sup>	This study

<b>Other plasmids</b>					
pBAD1K-ctr	pMD004		Control plasmid	p15A, Kan <sup>R</sup>	(51)
pBAD1K- <i>qrrS</i>	pMD099	<i>qrrS</i>	Inducible <i>qrrS</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
pBAD1K- <i>qrr4</i>	pMD176	<i>qrr4</i>	Inducible <i>qrr4</i> expression plasmid	p15A, Kan <sup>R</sup>	This study
pBAD1K- <i>qrrS*</i> (M1)	pMD103	<i>qrrS*</i> (G72C)	Inducible <i>qrrS*</i> (M1) expression plasmid	p15A, Kan <sup>R</sup>	This study
pCMW-1C- <i>mKate2</i>	pYH010	<i>mKate2</i>	Promoterless plasmid for transcriptional reporters	p15A, Cm <sup>R</sup>	(15)
pCMW-1C- <i>PqrrS::mKate2</i>	pAF012	<i>PqrrS::mKate2</i>	Transcriptional reporter for QrrS	p15A, Cm <sup>R</sup>	This study
pKAS32	pKAS32		Suicide plasmid for allelic exchange	R6K, Amp <sup>R</sup>	(36)
pKAS32- $\Delta$ <i>qrrS</i>	pAS005	up/downstream flanks of <i>qrrS</i>	Suicide plasmid for <i>qrrS</i> knock-out	R6K, Amp <sup>R</sup>	This study
pKAS32- $\Delta$ <i>vca0830</i>	pAF013	up/downstream flanks of <i>vca0830</i>	Suicide plasmid for <i>vca0830</i> knock-out	R6K, Amp <sup>R</sup>	This study
pKAS32- <i>vca0830::3XFLAG</i>	pMH075	<i>vca0830::3XFLAG</i>	Suicide plasmid <i>vca0830::3XFLAG</i> allelic replacement	R6K, Amp <sup>R</sup>	This study
pBB1	---	<i>luxCDABE</i> ( <i>V. harveyi</i> )	Bioluminescence assay	Tet <sup>R</sup>	(6)

**Table S3:** Strains used in this study

Strain	Relevant markers / genotype	Reference / Source
<b><i>V. cholerae</i></b>		
KPS-0014	C6706 wild-type	(53)
KPS-0358	C6706 $\Delta qrr1-4$	(54)
KPS-0995	C6706 <i>hfq::3XFLAG</i>	(34)
KPVC-12735	C6706 $\Delta qrrS$	This study
KPVC-12032	C6706 $\Delta vca0830$	This study
KPVC-12548	C6706 <i>vca0830::3XFLAG</i>	This study
<b><i>E. coli</i></b>		
Top 10	<i>F- mcrA</i> $\Delta(mrr-hsdRMS-mcrBC)$ $\phi 80lacZ\Delta M15$ $\Delta lacX74$ <i>nupG</i> <i>recA1</i> <i>araD139</i> $\Delta(ara-leu)7697$ <i>galE15</i> <i>galK16</i> <i>rpsL(StrR)</i> <i>endA1</i> $\lambda$ -	Invitrogen
S17 $\lambda$ pir	$\Delta lacU169$ ( $\Phi lacZ\Delta M15$ ), <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i> , <i>λpir</i>	New England Biolabs

**Table S4:** Oligonucleotides used in this study

Name	Sequence 5' to 3' (5' P denotes a 5' monophosphate)	Description
<b>Oligonucleotides for linearization of plasmids</b>		
KPO-0092	CCACACATTATACGAGCCGA	Plasmid construction (pEVS143)
KPO-0196	GGAGAAACAGTAGAGAGTTGCC	Plasmid construction (pBAD1K)
KPO-0267	TAATAGGCCTAGGATGCATATG	Plasmid construction (pKAS32)
KPO-0268	CGTTAACAACCGGTACCTCTA	Plasmid construction (pKAS32)
KPO-1023	GTTTTTCTAGAGGATCCGGTGATTGATTGAG	Plasmid construction (pEVS143)
KPO-1397	GATCCGGTGATTGATTGAGC	Plasmid construction (pEVS143 and pBAD1K)
KPO-1702	ATGCATGTGCTCAGTATCTCTATC	Plasmid construction (pXG10)
KPO-1703	GCTAGCGGATCCGCTGG	Plasmid construction (pXG10 and pXG30)
KPO-2591	GTCGACAGGCCTAGTTG	Plasmid construction (pCMW-1C- <i>mKate2</i> )
KPO-2592	GCATGCAAAAAGACCCCTC	Plasmid construction (pCMW-1C- <i>mKate2</i> )
KPO-4646	CCATTCGCCATTCAGGCTG	Plasmid construction (pXG30)
<b>Oligonucleotides for GFP reporter fusions (RIL-seq target validation)</b>		
KPO-1708	GAGATACTGAGCACATGCATAATTGATTTGGGACTGTTCC CAA	Plasmid construction (pNP058)
KPO-1709	GAGCCAGCGGATCCGCTAGCCAATTCGATAAGACGCGTCAC	Plasmid construction (pNP058)
KPO-1720	GAGATACTGAGCACATGCATAGGTTGTTATTAGCAATCCGCG ATAC	Plasmid construction (pNP064)
KPO-1721	GAGCCAGCGGATCCGCTAGCCAACGACAAAAGACCGACAGC AAG	Plasmid construction (pNP064)
KPO-2573	GAGATACTGAGCACATGCATAATCCCTGTCAGGTGTAAG	Plasmid construction (pMD092)
KPO-2574	GAGCCAGCGGATCCGCTAGCACCAGCACCTAACAGCAG	Plasmid construction (pMD092)
KPO-2779	GAGATACTGAGCACATGCATCGGAAAATATAATGCAAAAAGT GG	Plasmid construction (pMD161)

CHAPTER 5

KPO-2780	GAGCCAGCGGATCCGCTAGCGATTAAGTTATTAGAATTGCTGGG	Plasmid construction (pMD161)
KPO-3003	GTTTTTATGCATGATACTAATATAACGCATAACAATATC	Plasmid construction (pYH033)
KPO-3004	GTTTTTGCTAGCATGATGACGCTGTGCGCTTA	Plasmid construction (pYH033)
KPO-3005	GTTTTTATGCATGTACATGTGTAACCGATGGG	Plasmid construction (pYH034)
KPO-3006	GTTTTTGCTAGCAGCAATCCCCACTAGCAATC	Plasmid construction (pYH034)
KPO-3054	GTTTTTATGCATGCTCGTCACAGAACGAAATAC	Plasmid construction (pYH038)
KPO-3055	GTTTTTGCTAGCAAAACAGAGGGCAAGCAAACAAG	Plasmid construction (pYH038)
KPO-3795	GAGATACTGAGCACATGCATGTTTTGTTGGCATGGTCGC	Plasmid construction (pJR026)
KPO-3796	CCAGCGGATCCGCTAGCAATACGTTGCCGGTTTAGC	Plasmid construction (pJR026)
KPO-3813	GAGATACTGAGCACATGCATCTAAAAACCATCAGTCCCC	Plasmid construction (pJR029)
KPO-3814	CCAGCGGATCCGCTAGCACTATCCTTTTCAGGCATTTCG	Plasmid construction (pJR029)
KPO-4019	GAGATACTGAGCACATGCATGGCTCTTAGGTAAGAGTTGTT	Plasmid construction (pJR036)
KPO-4020	CCAGCGGATCCGCTAGCGAAACCAAACCTTTAGTTTCG	Plasmid construction (pJR036)
KPO-4056	GAGATACTGAGCACATGCATTTCTTCTGCGTTAAGCGCAA	Plasmid construction (pMH071)
KPO-4057	CCAGCGGATCCGCTAGCAGCCATTCCGGTGACCATC	Plasmid construction (pMH071)
KPO-4058	GAGATACTGAGCACATGCATATATTTGATGACATCCATTAACG	Plasmid construction (pMH058)
KPO-4059	CCAGCGGATCCGCTAGCATGCTCTTTGGCACGAACAC	Plasmid construction (pMH058)
KPO-4060	GAGATACTGAGCACATGCATACAAAATTAATGAGGCTACCTT	Plasmid construction (pMH056)
KPO-4061	CCAGCGGATCCGCTAGCTGCAACCACGCCTAGAAAG	Plasmid construction (pMH056)
KPO-4078	GAGATACTGAGCACATGCATCAAACAAAATGAAGGAGAAAGA	Plasmid construction (pMH059)
KPO-4079	CCAGCGGATCCGCTAGCATGCTCTTTGGCACGAACAC	Plasmid construction (pMH059)
KPO-4132	GAGATACTGAGCACATGCATATAATACGCATAATAAAACCCG	Plasmid construction (pJR040)
KPO-4133	CCAGCGGATCCGCTAGCCCAGAGGGTGTTCTTTTGC	Plasmid construction (pJR040)
KPO-4136	CCAGCGGATCCGCTAGCCATAGGGCTTCTCACTTTAAA	Plasmid construction (pMH066)
KPO-4137	GAGATACTGAGCACATGCATTTCGATGGTCACCTGAGTTG	Plasmid construction (pJR039)
KPO-4138	CCAGCGGATCCGCTAGCTGGGCAGTTTACCCGGG	Plasmid construction (pJR039)
KPO-4184	GAGATACTGAGCACATGCATGAATTTTAGGTAAGCCATTGG	Plasmid construction (pJR042)
KPO-4185	CCAGCGGATCCGCTAGCGAGTGCGCAGCCACTTAATG	Plasmid construction (pJR042)
KPO-4210	GAGATACTGAGCACATGCATAAAAAATTTCTCCCTAGCACAC	Plasmid construction (pMH063)
KPO-4211	CCAGCGGATCCGCTAGCCGCTGGTAGGAATCGGGTA	Plasmid construction (pMH063)
KPO-4212	GAGATACTGAGCACATGCATGTTTTATTTTCTAGCTTCAAATAAAG	Plasmid construction (pMH062)
KPO-4213	CCAGCGGATCCGCTAGCTGCAGCAGGACGAGTGTG	Plasmid construction (pMH062)
KPO-4276	GAGATACTGAGCACATGCATGAAAATGCCACTTTTTTCATAAAGTC	Plasmid construction (pMH060)

KPO-4277	CCAGCGGATCCGCTAGCTTGGCGGACCACGGTATAG	Plasmid construction (pMH060)
KPO-4469	GAGATACTGAGCACATGCATATTCCTATGCTGAGCGC	Plasmid construction (pJR043)
KPO-4470	CCAGCGGATCCGCTAGCATAAATTTGATAAATTGCAGCGG	Plasmid construction (pJR043)
KPO-4471	GAGATACTGAGCACATGCATAGGACTTACACCGTATCAAC	Plasmid construction (pJR044)
KPO-4472	CCAGCGGATCCGCTAGCTCGGTCTGCATTCAAAAATTG	Plasmid construction (pJR044)
KPO-4473	GAGATACTGAGCACATGCATCTGCAAGAGCTGAACCGG	Plasmid construction (pJR045)
KPO-4474	CCAGCGGATCCGCTAGCATCGACCAACGCGACACG	Plasmid construction (pJR045)
KPO-4630	CCAGCGGATCCGCTAGCGCTGCCTTTAGGGATAATAG	Plasmid construction (pMH069)
KPO-4642	GAGATACTGAGCACATGCATAAAAAACCTACTTATCTACGTC	Plasmid construction (pMH067)
KPO-4643	CCAGCGGATCCGCTAGCGGCAAAGGTGAGCAGAGG	Plasmid construction (pMH067)
KPO-4644	GAGATACTGAGCACATGCATTTGAATTTTGCAGTCAAACGAC	Plasmid construction (pMH064)
KPO-4645	CCAGCGGATCCGCTAGCGTTGCCATCAAATACACGTT	Plasmid construction (pMH064)
KPO-4647	CAGCCTGAATGGCGAATGGGCACTGGCAACGGGCATT	Plasmid construction (pMH068)
KPO-4648	CCAGCGGATCCGCTAGCGCGAACGGCTTGACCAAAG	Plasmid construction (pMH068)
KPO-4649	CAGCCTGAATGGCGAATGGGCGTATACCCTGCAAGAGAT	Plasmid construction (pMH065)
KPO-4650	CCAGCGGATCCGCTAGCGATCGCATCAAATTCCTTCGG	Plasmid construction (pMH065)
KPO-4651	CAGCCTGAATGGCGAATGGCACTGGCAGTTGCACGC	Plasmid construction (pMH066)
KPO-4929	CAGCCTGAATGGCGAATGGAAGTTGGTGGATGAAATCATT	Plasmid construction (pMH069)
KPO-4935	CAGCCTGAATGGCGAATGGCCACTCAGCGCTTTCCGT	Plasmid construction (pMH072)
KPO-4936	CCAGCGGATCCGCTAGCGGTGCTTCCCGCGCTTTG	Plasmid construction (pMH072)
KPO-4937	GAGATACTGAGCACATGCATGTCCGTATAGTGACACAGA	Plasmid construction (pMH073)
KPO-4938	CCAGCGGATCCGCTAGCAATGGTCTGCAAATCTAAATTAAC	Plasmid construction (pMH073)
KPO-5187	CAGCCTGAATGGCGAATGGCGTTGGTTGATTGATAGCCG	Plasmid construction (pKT003)
KPO-5188	CCAGCGGATCCGCTAGCTGCATTGGCGCCAGCCAA	Plasmid construction (pKT003)
KPO-5189	CAGCCTGAATGGCGAATGGCGTATTCCGCCGCGACAT	Plasmid construction (pKT002)
KPO-5190	CCAGCGGATCCGCTAGCCCATGGGTCAATTATCGCGG	Plasmid construction (pKT002)
KPO-5191	GAGATACTGAGCACATGCATGCATTACCATCAGTAAGCGC	Plasmid construction (pKT001)
KPO-5192	CCAGCGGATCCGCTAGCTTGAATCAGTTGATAAACTTTTTTC	Plasmid construction (pKT001)
KPO-5208	CAGCCTGAATGGCGAATGGCCCATCACGGGCATTACC	Plasmid construction (pKT004)
KPO-5209	CCAGCGGATCCGCTAGCATACGCTTTACTGACTTGCTG	Plasmid construction (pKT004)
KPO-5367	CCAGCAGCGGAGCCAGCGGATCCGCTAGCCAATATATTGCG ATCTATAC	Plasmid construction (pKT005)
KPO-5369	GAGCCAGCGGATCCGCTAGCCGTAGCGATAGGTAGCATC	Plasmid construction (pKT008)
KPO-5371	CCAGCAGCGGAGCCAGCGGATCCGCTAGCAAACGAGTTGTT GGTAACC	Plasmid construction (pKT007)

CHAPTER 5

KPO-5408	GAGATACTGAGCACATGCATAAGGAAGTAATAAGGTGGAATA	Plasmid construction (pKT005)
KPO-5409	GAGATACTGAGCACATGCATGTTATTCGCTCCCTTTTTTATG	Plasmid construction (pKT008)
KPO-5410	GAGATACTGAGCACATGCATGCTATCTAGCAGAAGAGGAG	Plasmid construction (pKT007)
KPO-5411	GAGATACTGAGCACATGCATCAAATTCGATAGAACAACCCT	Plasmid construction (pKT006)
KPO-5412	CCAGCGGATCCGCTAGCTGGAATCATGCCACACGAT	Plasmid construction (pKT006)
<b>Oligonucleotides for sRNA expression plasmids (RIL-seq target validation)</b>		
KPO-1005	P-GTCATCTCGTTAGTCATTACGA	Plasmid construction (pNP004)
KPO-1006	GTTTTTCTAGACACTGACAAACCGGTGTTGG	Plasmid construction (pNP004)
KPO-1076	P-GCGTAGGGTACAGAGGTAA	Plasmid construction (pAS001)
KPO-1077	GTTTTTCTAGAAGTGCCAACGTGGAATAGC	Plasmid construction (pAS001)
KPO-1084	P-GCAACGGCGGCCCTGAACGG	Plasmid construction (pRH006)
KPO-1085	GTTTTTCTAGAAGCTCAGTATTTACTGGTTGGA	Plasmid construction (pRH006)
KPO-1090	P-TGACCCTTCTAAGCCGAGG	Plasmid construction (pRH003)
KPO-1091	GTTTTTCTAGACCACGAAAGCCAAGATGCT	Plasmid construction (pRH003)
KPO-1092	P-ACAAAGTATCACAAAAATCAGGG	Plasmid construction (pRH002)
KPO-1093	GTTTTTCTAGAAAAGCAGTGAAAATAGCGGG	Plasmid construction (pRH002)
KPO-1858	TCGGCTCGTATAATGTGTGGGCTAGCGAAAATATAATCATA AAC	Plasmid construction (pSG001)
KPO-1859	CTCAATCAATCACCGGATCCGCTTTGATTGAGCAGACGTTG	Plasmid construction (pSG001)
KPO-2565	GCTCAATCAATCACCGGATCATTGAAGTGAGTGATGGTAA TAG	Plasmid construction (pMD104)
KPO-2570	TCGGCTCGTATAATGTGTGGGTTGAAAGGACATCCCTCC	Plasmid construction (pMD104)
KPO-3965	GGCTCGTATAATGTGTGGGTACCCCTGATAATTCGTATC	Plasmid construction (pJR035)
KPO-3966	GCTCAATCAATCACCGGATCCTACACAGGGATTAATCTC	Plasmid construction (pJR035)
KPO-4062	GGCTCGTATAATGTGTGGATTTTTCTTGGGCTTCCCC	Plasmid construction (pMH057)
KPO-4063	GCTCAATCAATCACCGGATCGGCGGGTTTCTCATTGTG	Plasmid construction (pMH057)
<b>Oligonucleotides for construction of other plasmids</b>		
KPO-1301	CCAGTTAACTTGAGATGAAAATGGTGCAAGGTTGAATTTTTGT TAGTG	Plasmid construction (pAS005)
KPO-1302	CATTTTCATCTCAAGTTAACTGG	Plasmid construction (pAS005)
KPO-1304	GTTTTTGGTACCGAATGCGTTGTAACCTATGAAC	Plasmid construction (pAS005)
KPO-1305	GTTTTTCTAGGTACGGCATAAGTCATGACTCG	Plasmid construction (pAS005)
KPO-2558	CGCAACTCTCTACTGTTTCTCCTGAATAATCAAAGACGAGGC	Plasmid construction (pMD099)
KPO-2559	GCTCAATCAATCACCGGATCGAACAGCCAGTTAACTTGAG	Plasmid construction (pMD099)
KPO-3676	GAAGGGTCTTTTTGCATGCCTACTTGTACAGCGGCTTTAT	Plasmid construction (pAF012)
KPO-3677	CAACTAGGCCTGTGCGACTATTCAATCATTAGTTTAGGTGCAA	Plasmid construction (pAF012)

KPO-3741	TAGAGGTACCGGTTGTTAACGGTCTCTGCACTCGCCAG	Plasmid construction (pAF013)
KPO-3742	ACCGATAAATGTGCCAAGATAAC	Plasmid construction (pAF013)
KPO-3743	TCTTGGCACATTTATCGGTGATGAAAATGGGTGAAAAAGGAAAG	Plasmid construction (pAF013)
KPO-3744	CATATGCATCCTAGGCCTATTAGAACATCTGACAAAAATCCTCTC	Plasmid construction (pAF013)
KPO-3749	CAACTCATGCCTAGGGTTATTAACAAAAATCAAAAAC	Plasmid construction (pMD103)
KPO-3750	TAACCCTAGGCATGAGTTGTCAGTCAATAAACAG	Plasmid construction (pMD103)
KPO-3779	CGCAACTCTCTACTGTTTCTCCGACCCTTCTAAGCCGAGG	Plasmid construction (pMD176)
KPO-3780	GCTCAATCAATCACCGGATCCCACGAAAGCCAAGATGCTAT	Plasmid construction (pMD176)
<b>Oligonucleotides for Northern blot probing</b>		
KPO-0063	CGTCTATAAGTGTGAACAATGGTG	Qrr4 oligoprobe
KPO-0243	TTCGTTTCACTTCTGAGTTCCGG	5S oligoprobe
KPO-3749	GACCCTTTCCTTTGTTGCTC	QrrS oligoprobe

## REFERENCES

1. Schauder S & Bassler BL (2001) The languages of bacteria. *Genes Dev* 15(12):1468-1480.
2. Pappenfort K & Bassler BL (2016) Quorum sensing signal-response systems in Gram-negative bacteria. *Nat Rev Microbiol* 14(9):576-588.
3. Rutherford ST & Bassler BL (2012) Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harb Perspect Med* 2(11).
4. Miller MB & Bassler BL (2001) Quorum sensing in bacteria. *Annu Rev Microbiol* 55:165-199.
5. Ng WL & Bassler BL (2009) Bacterial quorum-sensing network architectures. *Annu Rev Genet* 43:197-222.
6. Miller MB, Skorupski K, Lenz DH, Taylor RK, & Bassler BL (2002) Parallel quorum sensing systems converge to regulate virulence in *Vibrio cholerae*. *Cell* 110(3):303-314.
7. Chen X, *et al.* (2002) Structural identification of a bacterial quorum-sensing signal containing boron. *Nature* 415(6871):545-549.
8. Higgins DA, *et al.* (2007) The major *Vibrio cholerae* autoinducer and its role in virulence factor production. *Nature* 450(7171):883-886.
9. Pappenfort K, *et al.* (2017) A *Vibrio cholerae* autoinducer-receptor pair that controls biofilm formation. *Nat Chem Biol* 13(5):551-557.
10. Lenz DH, *et al.* (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* 118(1):69-82.
11. Shao Y & Bassler BL (2012) Quorum-sensing non-coding small RNAs use unique pairing regions to differentially control mRNA targets. *Mol Microbiol* 83(3):599-611.
12. Rutherford ST, van Kessel JC, Shao Y, & Bassler BL (2011) AphA and LuxR/HapR reciprocally control quorum sensing in *vibrios*. *Genes Dev* 25(4):397-408.
13. Bardill JP, Zhao X, & Hammer BK (2011) The *Vibrio cholerae* quorum sensing response is mediated by Hfq-dependent sRNA/mRNA base pairing interactions. *Mol Microbiol* 80(5):1381-1394.
14. Pappenfort K, Forstner KU, Cong JP, Sharma CM, & Bassler BL (2015) Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proc Natl Acad Sci U S A* 112(7):E766-775.
15. Herzog R, Peschek N, Frohlich KS, Schumacher K, & Pappenfort K (2019) Three autoinducer molecules act in concert to control virulence gene expression in *Vibrio cholerae*. *Nucleic Acids Res* 47(6):3171-3183.
16. Updegrove TB, Zhang A, & Storz G (2016) Hfq: the flexible RNA matchmaker. *Curr Opin Microbiol* 30:133-138.
17. Santiago-Frangos A & Woodson SA (2018) Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip Rev RNA* 9(4):e1475.
18. Huber M, Frohlich KS, Radmer J, & Pappenfort K (2020) Switching fatty acid metabolism by an RNA-controlled feed forward loop. *Proc Natl Acad Sci U S A* 117(14):8044-8054.
19. Zhang A, *et al.* (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol* 50(4):1111-1124.
20. Sittka A, *et al.* (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* 4(8):e1000163.
21. Chao Y, Pappenfort K, Reinhardt R, Sharma CM, & Vogel J (2012) An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* 31(20):4005-4019.

22. Bilusic I, Popitsch N, Rescheneder P, Schroeder R, & Lybecker M (2014) Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. *RNA Biol* 11(5):641-654.
23. Chao Y & Vogel J (2010) The role of Hfq in bacterial pathogens. *Curr Opin Microbiol* 13(1):24-33.
24. Ding Y, Davis BM, & Waldor MK (2004) Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol Microbiol* 53(1):345-354.
25. Barquist L & Vogel J (2015) Accelerating Discovery and Functional Analysis of Small RNAs with New Technologies. *Annu Rev Genet* 49:367-394.
26. Adams PP & Storz G (2020) Prevalence of small base-pairing RNAs derived from diverse genomic loci. *Biochim Biophys Acta Gene Regul Mech* 1863(7):194524.
27. Melamed S, *et al.* (2018) Mapping the small RNA interactome in bacteria using RIL-seq. *Nat Protoc* 13(1):1-33.
28. Melamed S, *et al.* (2016) Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol Cell* 63(5):884-897.
29. Figueroa-Bossi N & Bossi L (2018) Sponges and Predators in the Small RNA World. *Microbiol Spectr* 6(4).
30. Corcoran CP, *et al.* (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* 84(3):428-445.
31. Frohlich KS, Haneke K, Papenfort K, & Vogel J (2016) The target spectrum of SdsR small RNA in *Salmonella*. *Nucleic Acids Res* 44(21):10406-10422.
32. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.
33. Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res* 37(20):6984-6990.
34. Peschek N, Hoyos M, Herzog R, Forstner KU, & Papenfort K (2019) A conserved RNA seed-pairing domain directs small RNA-mediated stress resistance in enterobacteria. *EMBO J* 38(16):e101650.
35. Dunn AK, Millikan DS, Adin DM, Bose JL, & Stabb EV (2006) New rfp- and pES213-derived tools for analyzing symbiotic *Vibrio fischeri* reveal patterns of infection and lux expression in situ. *Appl Environ Microbiol* 72(1):802-810.
36. Skorupski K & Taylor RK (1996) Positive selection vectors for allelic exchange. *Gene* 169(1):47-52.
37. Heidelberg JF, *et al.* (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406(6795):477-483.
38. Rehmsmeier M, Steffen P, Hochsmann M, & Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. *RNA* 10(10):1507-1517.
39. Maddocks SE & Oyston PCF (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* 154(Pt 12):3609-3623.
40. Bassler BL, Wright M, & Silverman MR (1994) Multiple signalling systems controlling expression of luminescence in *Vibrio harveyi*: sequence and function of genes encoding a second sensory pathway. *Mol Microbiol* 13(2):273-286.
41. Waters SA, *et al.* (2017) Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *EMBO J* 36(3):374-387.
42. Iosub IA, *et al.* (2020) Hfq CLASH uncovers sRNA-target interaction networks linked to nutrient availability adaptation. *Elife* 9.
43. Faigenbaum-Romm R, *et al.* (2020) Hierarchy in Hfq Chaperon Occupancy of Small RNA Targets Plays a Major Role in Their Regulation. *Cell Rep* 30(9):3127-3138 e3126.
44. Feng L, *et al.* (2015) A Qrr noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell* 160(1-2):228-240.

## CHAPTER 5

45. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16(22):10881-10890.
46. Song T, *et al.* (2008) A new *Vibrio cholerae* sRNA modulates colonization and affects release of outer membrane vesicles. *Mol Microbiol* 70(1):100-111.
47. Song T, Sabharwal D, & Wai SN (2010) VrrA mediates Hfq-dependent regulation of OmpT synthesis in *Vibrio cholerae*. *J Mol Biol* 400(4):682-688.
48. Yamamoto S, *et al.* (2011) Identification of a chitin-induced small RNA that regulates translation of the *tfoX* gene, encoding a positive regulator of natural competence in *Vibrio cholerae*. *J Bacteriol* 193(8):1953-1965.
49. Davis BM & Waldor MK (2007) RNase E-dependent processing stabilizes MicX, a *Vibrio cholerae* sRNA. *Mol Microbiol* 65(2):373-385.
50. Hammer BK & Bassler BL (2007) Regulatory small RNAs circumvent the conventional quorum sensing pathway in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 104(27):11145-11149.
51. Hoyos M, Huber M, Forstner KU, & Papenfort K (2020) Gene autoregulation by 3' UTR-derived bacterial small RNAs. *Elife* 9.
52. Waters CM & Bassler BL (2006) The *Vibrio harveyi* quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers. *Genes Dev* 20(19):2754-2767.
53. Thelin KH & Taylor RK (1996) Toxin-coregulated pilus, but not mannose-sensitive hemagglutinin, is required for colonization by *Vibrio cholerae* O1 El Tor biotype and O139 strains. *Infect Immun* 64(7):2853-2856.
54. Svenningsen SL, Tu KC, & Bassler BL (2009) Gene dosage compensation calibrates four regulatory RNAs to control *Vibrio cholerae* quorum sensing. *EMBO J* 28(4):429-439.

## 6 Discussion

### 6.1 Detecting sRNAs and their targets with high-throughput approaches

High-throughput sequencing based approaches are one of the major technical advances in molecular biology in the last decades. In the field of small regulatory RNAs, these techniques have provided enormous opportunities, and led to a constant surge of newly discovered RNA regulators in a variety of microorganisms (3, 204). Nevertheless, high-throughput methods also have limitations and possible biases, which should also be addressed critically. In this thesis, RIP-seq (RNA co-immunoprecipitation followed by RNA sequencing) and RIL-seq (RNA interaction by ligation and sequencing) have been used to investigate Hfq-binding sRNAs and their targets and will be discussed below.

#### 6.1.1 Global detection of Hfq-binding sRNAs via co-immunoprecipitation

In chapters 2 and 4, RIP-seq was performed to globally detect association of RNAs with the RNA chaperone Hfq in *V. cholerae* and *P. laumondii*, respectively. RIP-seq has been shown previously to be a very reliable method for this purpose (29, 198, 205), and Western blot analysis of immunoprecipitated Hfq as well as Northern blot analysis of known Hfq-binding partners in *V. cholerae* have confirmed the specificity of our experimental approach (chapter 2, Fig. 1 A and B). In line with the expectation that Hfq binds specifically to sRNAs and mRNAs, no enrichment of tRNA or rRNA has been observed in our data. Further, all Hfq-binding sRNAs in *V. cholerae* that have been characterized in detail previously (39, 62, 80, 191, 192, 195, 206, 207) were recovered in our experiments (chapter 2, Table S2). The vast majority of additional candidate sRNAs that showed reduced transcript levels in an *hfq* mutant in a previous study (39) were also enriched in the Hfq co-immunoprecipitation (chapter 2, Table S2). For *P. laumondii*, there were no Hfq-binding sRNAs previously studied in detail, which could have served as a positive control, but several conserved sRNAs known as RNA-ligands of Hfq in *E. coli* and *Salmonella* (e.g. RyhB, RybB, GcvB, RprA (25, 56, 58, 64)) were significantly enriched in our RIP-seq data (chapter 4, Table S4).

## 6 DISCUSSION

RIP-seq clearly captured RNA-Hfq interactions reliably at a large scale, providing valuable insights into Hfq-mediated regulation at a given condition. However, a technical bias which cannot be completely eliminated might be that different steps of the procedure such as cell lysis and multiple rounds of washing could favor more stable sRNAs to be detected. Furthermore, despite the reliability of the approach, validation on an individual basis is nevertheless a fundamental step before studying sRNAs in more detail. The gold standard in this field to prove an independent sRNA transcript is Northern blot analysis. Northern blot analysis in general not only provides information about the expression and, in the case of co-immunoprecipitation, the specific enrichment in the Hfq pull-down, but also information about the length of the transcript as well as possible processing intermediates and degradation products. Therefore, extensive Northern blot analysis has been performed in this study (chapter 2, Figs. 1E and S1B; chapter 4, Ext. Data Figs. 1 and 2). Electrophoretic mobility shift assays were carried out additionally in chapter 2 (Fig. S2) in order to show specific Hfq binding of individual sRNA transcripts *in vitro*.

In addition to detecting Hfq binding of previously identified sRNAs or predicted candidates, the RIP-seq data set in chapter 2 served to detect putative novel sRNAs in *V. cholerae*, which had not been predicted before by other approaches, e.g. differential RNA-seq. To this end, regions with a strong enrichment of reads in the Hfq::3XFLAG strain were examined for characteristic features of Hfq-binding sRNAs. As this approach does not rely on transcriptional start sites, sRNAs deriving from 3' UTRs of coding sequences which do not have their own promoter can be detected by this approach. The fact that RIP-seq and related methods have been used more and more to detect novel sRNAs could explain why in the last years an increasing number of sRNAs from 3' UTRs have been discovered (30, 171), while most of the early identified sRNAs derive from IGRs, since searches here often screened for promoter elements and transcriptional start sites and focused on intergenic regions (24).

The advantage of detecting sRNAs via their binding to Hfq is at the same time a limitation, as sRNAs associating with other RNA-binding proteins (RBPs) cannot be found. However, it has been shown that RIP-seq and related methods, such as CLIP-seq, can also be applied successfully to other RBPs, in particular ProQ and CsrA ((136, 208) and unpublished data in the lab). Moreover, a method has recently been developed that facilitates the unbiased identification of putative new RBPs. The underlying principle of this technique, called Grad-seq (gradient profiling by sequencing), is to partition sRNAs by their biochemical properties providing hints to proteins binding to certain clusters of sRNAs (135).

### 6.1.2 Identification and validation of target RNAs

Being a well-established and reliable method to detect Hfq association of RNA molecules, RIP-seq, however, does not provide any information about base-pairing partners of these RNAs. A widely used approach to identify targets regulated by sRNAs is temporary overexpression of the sRNA of interest followed by whole transcriptome sequencing (e.g. chapter 2, Fig. S4A and Table 1; chapter 3, Fig. 3A and Fig. 3 – suppl. 1A). This allows to monitor changes in transcript levels on a global scale. Short overexpression instead of constitutive overexpression is used here in order to detect differences in target mRNA levels which are a direct consequence of sRNA-mRNA interaction and not indirect effects involving additional factors. However, this also means that there is a requirement for a change in transcript levels to identify a target, which could mask the detection of some targets. Another disadvantage of this set-up is the artificial situation of the strong overexpression of the sRNA, which could bias the data. Further, the approach is rather labor-intense as it has to be performed individually for each sRNA of interest.

Given these drawbacks, several techniques have recently been developed to reveal information about sRNA-target interactions by other means and on a more global scale. RIL-seq as well as CLASH are methods based on co-immunoprecipitation but including RNA ligation to unravel physical interactions of RNA molecules mediated by an RNA-binding protein. Although the two methods have some differences in the experimental procedures, e.g. denaturing of the Hfq hexamer in CLASH vs. keeping the hexamer intact in RIL-seq, the key point of both methods is to capture RNAs simultaneously bound to the protein (168, 171). RIL-seq has been successfully established in *E. coli*, and apart from Hfq has also been applied to ProQ ((68, 126, 168) and own unpublished data for *V. cholerae*). CLASH, originally developed with eukaryotic Argonaute proteins (169), has been performed in *E. coli* with the ribonuclease RNase E as bait (170) as well as with Hfq (171). A third method to identify targets of sRNAs is MAPS (MS2-affinity purification coupled with RNA sequencing) (209, 210). In contrast to RIL-seq and CLASH, in this approach no RBP is involved; instead, the sRNA of interest is tagged with the MS2 RNA aptamer, allowing the identification of specific targets by affinity chromatography and sequencing.

In this thesis (chapter 5), RIL-seq analysis has been applied to *V. cholerae* Hfq. To our knowledge, this is the first attempt to study the global RNA-RNA interactome of this model organism as well as one of the first studies to transfer the method developed in *E. coli* to another species. We identified approximately 3,000 potential interactions at two different growth conditions (chapter 5, Fig. 1 A and B), which also recovered several previously described sRNA targets (chapter 5, Table S1). Nevertheless, some known interactions in *V. cholerae* do not appear in our data set (e.g. MicV-*ompT* (62)). This is not surprising, given

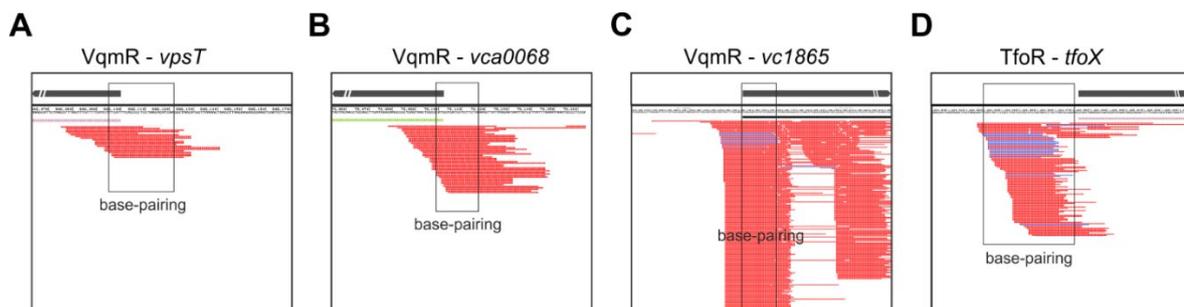
## 6 DISCUSSION

that in the original RIL-seq study in *E. coli* only approximately 56% of the known sRNA-target pairs have been recovered under the tested conditions (126). There are several possible reasons for this: First, a technical limitation of RIL-seq might be that sequence and structure of RNAs could influence the ligation process. Possibly, this explains why some Hfq-binding sRNAs and their targets do not appear with significant chimeras in the RIL-seq data. Second, the RIL-seq data are a snapshot of the RNA-RNA interactome at a given condition. Here, we performed the experiments at two growth phases (low and high cell densities). This should cover a broad range of sRNAs, however, some sRNAs or their targets might be expressed only under very specific conditions and therefore might not be detected in our experiments.

To validate sRNA-mRNA interactions predicted by RIL-seq, we used a well-established two plasmid system based on translational GFP reporter fusions (211). Although a significant number of mRNA targets of several sRNAs could be validated with this approach (chapter 5, Fig. 1C), about one third of predicted targets could not be confirmed in these experiments. Different aspects come into play here: First, a certain fraction might be explained by technical false positives occurring in the RIL-seq analysis. Although the pull-down of chimeric fragments mediated by Hfq is very specific, given the negligible number of chimeras in the untagged control strain (7 at OD<sub>600</sub> of 0.2 and 14 at OD<sub>600</sub> of 2.0) in contrast to the about 3,000 chimeras in the Hfq::3XFLAG strain at both growth conditions, the detection of false positives, possibly caused by ligation of RNAs randomly in physical proximity, cannot be avoided in such large scale approaches and might account for a certain percentage of predicted interactions that cannot be recovered by other methods. To address this problem, the odds ratio has been calculated, providing information about over-representation of a given RNA-RNA pair above random pairing (126). Further, and maybe more interesting, there might be cases where the interaction is true, but there is no regulation. This seems to be a relevant scenario, which might have been overlooked in the past. Indeed, *Faigenbaum-Romm* (125) observed a high percentage of so called “unaffected targets” which constitute reproducible interactions, but do not involve regulation of expression. Related to this, we also found non-canonical interactions far deeper in the coding sequence than usually described for target regulation. For these “unaffected targets”, translational GFP reporter fusions are not appropriate, and other means have to be used in order to test, validate and understand these interactions. To have a measure of how likely an interaction also results in a change of expression level of the target, the normalized odds ratio is determined, which is based on the observation that interactions with a higher enrichment more likely involve target regulation (normalized odds ratio = odds ratio multiplied by the relative enrichment of the two RNAs) (126).

In general, the results point towards a broader concept of “sponge interactions” in RNA-based regulation, where targets are not only regulated by sRNAs, but at the same time play important roles acting as sponges of regulators. The fact that these connections can be detected with RIL-seq and similar methods opens the door for future studies on these types of interactions. Particularly interesting are sRNA-sRNA pairs where one sRNA acts as a sponge of another, thereby titrating it away from its mRNA targets. Identifying these interactions, as e.g. the QrrS sRNA investigated in detail in this thesis (chapter 5, and discussed in section 6.2) or the interaction between the well-known sRNAs ArcZ and CyaR, identified when applying CLASH to *E. coli* Hfq (171), can provide important new insights into complex regulatory networks and foster a deeper understanding of sRNA-based gene regulation.

Another benefit of RIL-seq and CLASH is that they provide information about Hfq binding sites and RNA-RNA pairing regions. Although it cannot be generally assumed that the cDNA reads of the chimeric fragments represent the exact base-pairing site, the position of the reads still hints to a certain region which might have been protected by the Hfq protein during the procedure, and therefore was not subject to RNase digestion and consequently was available for cDNA preparation. A comparison of the previously predicted base-pairing regions between the extensively studied *V. cholerae* VqmR sRNA and three of its validated mRNA targets (39) with the positions of chimeras found by RIL-seq indeed shows a large overlap (Fig. 6.1 A-C). Similarly, chimeric fragments between the TfoR sRNA and the *tfoX* mRNA widely overlap with the previously identified region of base-pairing (207) (Fig. 6.1 D).



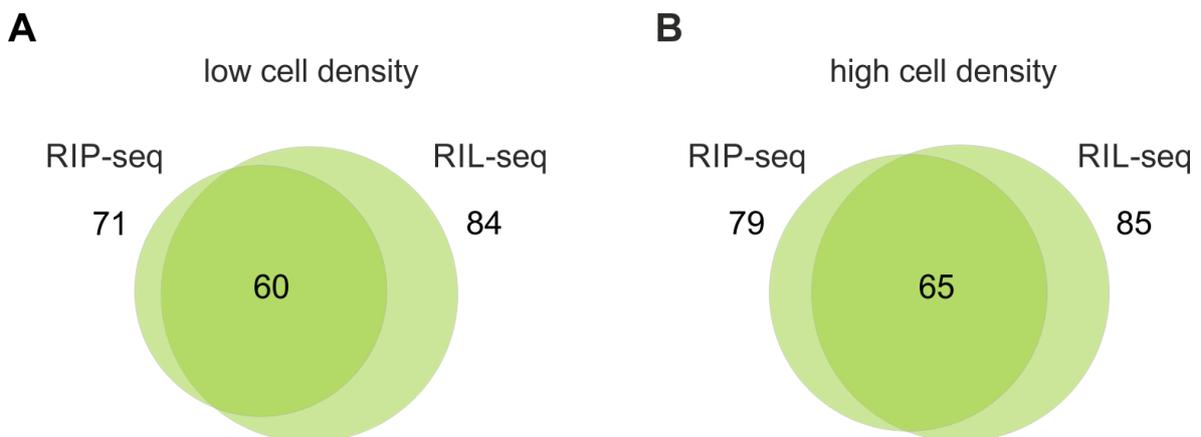
**Figure 6.1: Position of chimeric fragments relative to base-pairing regions**

Screenshots visualizing the positions of sRNA chimeras with known mRNA targets, as well as previously predicted base-pairing interactions between the sRNAs and the respective targets (39, 207). Target genes are shown in gray, chimeric fragments are shown in red. Base-pairing regions are indicated by boxes. Data has been displayed with the UCSC Genome Browser (<http://genome.ucsc.edu/>).

### 6.1.3 Integration of multiple data sets

Given that more and more high-throughput data sets are generated, a major challenge is to compare and utilize these data sets in a way to learn the most from them. Even if based on slightly different experimental set-ups and conditions, integrating different data sets can be valuable to deal with technical biases specific to each approach and consequently better overall understanding of RNA-RNA interactions. Although overlap of data generated with different techniques might be rather low, as for example when comparing the *E. coli* RIL-seq and CLASH data (126, 171), detailed comparison still might help for a deeper understanding of the data, and data sets can complement each other.

In this thesis, two large data sets have been produced providing information about association of RNA-ligands with Hfq in *V. cholerae*. Although in the RIP-seq data (chapter 2) enrichment of individual transcripts in the Hfq::3XFLAG strain is calculated in comparison to the untagged control whereas in the RIL-seq analysis (chapter 5) significant chimeras found on Hfq are presented, it might be interesting to compare the two data sets. Out of the enriched sRNAs in the co-immunoprecipitation, indeed 85% and 82%, respectively, are also found with significant chimeras in the RIL-seq experiment at the two growth conditions. Vice versa, 71% and 76% of the sRNAs detected to form chimeras were also enriched in the RIP-seq experiments (Fig. 6.2 A and B). This relatively high degree of overlap, despite the different procedures and analyses, supports that these are true Hfq-binding sRNAs. Of note, there might be sRNAs bound to Hfq in the RIL-seq experiment which did not engage base-pairing and therefore would not appear in the analysis. If these were additionally considered, the overlap might be even higher.



**Figure 6.2: Comparison of RIP-seq and RIL-seq datasets**

Venn-diagrams visualizing Hfq-bound sRNAs in *V. cholerae* RIP-seq and RIL-seq experiments at low and high cell densities.

In addition to experimental data sets, bioinformatic predictions are another source to reveal sRNA-target pairs. Based on thermodynamic models and including various parameters, such as sequence accessibility, several free online tools are available for target prediction. The CopraRNA algorithm (Comparative prediction algorithm for small RNA targets) (212), the leading method in this field, additionally takes into account conservation, thereby increasing the sensitivity and specificity of the predictions (212, 213). When looking at validated sRNA-target interactions in *E. coli*, according to *Melamed et al.* (126), the two approaches, the experimental RIL-seq and the bioinformatics tool CopraRNA, perform almost equally well. However, due to the completely distinct approach on predicting targets and different strength and weaknesses, the methods should be regarded as complementary rather than exchangeable (213). For instance, comparison of the targets of the RyhB sRNA in *E. coli* showed only a minor overlap of targets predicted by both RIL-seq and CopraRNA (213). Similar outcomes were obtained when comparing Spot 42 targets predicted with different methods (78).

In the RIL-seq data set for *V. cholerae* Hfq (chapter 5), Vcr017 was one of the sRNAs for which many interactions with different mRNA targets were predicted. Therefore, I performed CopraRNA analysis for Vcr017 as a representative example and compared predicted targets of both approaches. Out of the Top 20 hits in CopraRNA, 10 are also detected in the RIL-seq data set. However, several of them have only few numbers of chimeras and would therefore not appear among the Top hits in RIL-seq (Fig. 6.3 A). Conversely, when taking the 20 putative mRNA targets with the highest number of chimeras (Top 10 of each growth condition) in RIL-seq, only 4 are among the Top 20 hits in CopraRNA (Fig. 6.3 B). This shows one more time that various methods should be considered to comprehensively study RNA-RNA interactions.

## 6 DISCUSSION

A CopraRNA Top 20				B RIL-seq Top 20			
#	Gene	Description	Number of chimeras at LCD / HCD	#	Gene	Description	Number of chimeras
1	vc0619	peptide ABC transporter permease	187 / 51	1	vc2217	beta-N-acetylhexosaminidase	2,315
2	vc2217	beta-N-acetylhexosaminidase	2,315 / 305	2	vc0910/vc0911	PTS system trehalose-specific transporter subunits / trehalose-6-phosphate hydrolase	1,956
3	vc2589	50S ribosomal protein		3	vc2348	phosphopentomutase	919
4	vca0248	L-ascorbate 6-phosphate lactonase		4	vc2669/vc2668	5-carboxymethyl-2-hydroxyruconate delta-isomerase / hypothetical protein	715
5	vca0919	hypothetical protein	368 / 664	5	vc0328/vc0329	DNA-directed RNA polymerase subunit beta	646
6	vc2364	bifunctional aspartokinase / homoserine dehydrogenase ( <i>thrA</i> )	< 50	6	vc2271	riboflavin-specific deaminase	494
7	vca1046	mannitol-1-phosphate 5-dehydrogenase	< 50	7	vc1866	formate acetyltransferase	453
8	vc0153	hypothetical protein		8	vc2096/vc2095	replication initiation regulator SeqA / phosphoglucomutase	448
9	vc2668	hypothetical protein	< 50	9	vca0919	hypothetical protein	368
10	vc0070	hypothetical protein		10	vc0439/vc0440	hypothetical protein / dihydrofolate reductase	348
11	vc1434	fumarate / nitrate reduction transcriptional regulator	- / 209	1	vc0328/vc0329	DNA-directed RNA polymerase subunit beta	1,854
12	vca0944	maltose transporter membrane protein		2	vc2469	L-aspartate oxidase	1,218
13	vc0815	hypothetical protein		3	vc2763/vc2762	ATP synthase F0F1 subunit epsilon / glucosamine-1-phosphate acetyltransferase	992
14	vc1315	sensor histidine kinase	- / 59	4	vc2217	beta-N-acetylhexosaminidase	948
15	vc0420	hypothetical protein		5	vc0671	dinucleoside polyphosphate hydrolase	907
16	vca0957	malate synthase		6	vc0553/vc0554	hypothetical protein/insulinase family protease	756
17	vc2583	30S ribosomal protein S14 ( <i>rpsN</i> )	- / 110	7	vc2348	phosphopentomutase	720
18	vc0911	trehalose-6-phosphate hydrolase	1,956 / 117	8	vca0919	hypothetical protein	664
19	vca0083	multidrug resistance protein D		9	vca0156/vca0155	monovalent cation/H+ antiporter subunit / NADH dehydrogenase	657
20	vc0612	cellobiose / cellodextrin phosphorylase		10	vc2669/vc2668	5-carboxymethyl-2-hydroxyruconate delta-isomerase / hypothetical protein	642

**Figure 6.3: Putative targets of Vcr017 predicted by CopraRNA and / or RIL-seq**

**A)** Top 20 target predictions by CopraRNA. Putative targets recovered by RIL-seq are marked in blue. **B)** Top 20 (Top 10 at low and high cell densities, respectively) chimeric interactions identified by RIL-seq. Overlap with the Top 20 of CopraRNA predictions is marked in orange. Description is based on the annotation at KEGG (<https://www.genome.jp/kegg>).

To better understand functions of sRNAs and cellular RNA networks, it might further be interesting to compare the available data sets among species. Spot 42 is one of the best characterized sRNAs in *Enterobacteria* and regulates a large set of targets, most of them involved in carbohydrate metabolism (reviewed in (78)). Although the targets are not generally conserved across bacterial species (discussed in section 6.4), it is exciting to compare targets discovered in the *E. coli* RIL-seq and the *V. cholerae* RIL-seq data. Indeed, there are several overlapping targets, as for example *sdhC* (encoding succinate dehydrogenase), *galK* (encoding galactokinase) or *glcA* (encoding citrate synthase) (126), underlining the conserved function of the sRNA in carbohydrate metabolism and the reliability of the experiments. Apart from these, among the predicted targets in *V. cholerae*, there are several others involved in carbohydrate metabolism, e.g. *prpC* (encoding methylcitrate synthase) or *treB* (encoding trehalose transporter).

In general, to compare multiple data sets and profit optimally from the large amount of data generated by high-throughput sequencing experiments, a crucial challenge for the future will be to make these available in a user-friendly way for the whole research

community. For the best studied model organisms, such as *Salmonella*, there are already examples, such as the SalComMac database (<http://tinyurl.com/SalComMac>), visualizing comprehensive transcriptomic data. Improving the accessibility of large data resources, including RIP-seq, RIL-seq and similar data, can be a big chance for future studies on bacterial RNA-based gene expression regulation.

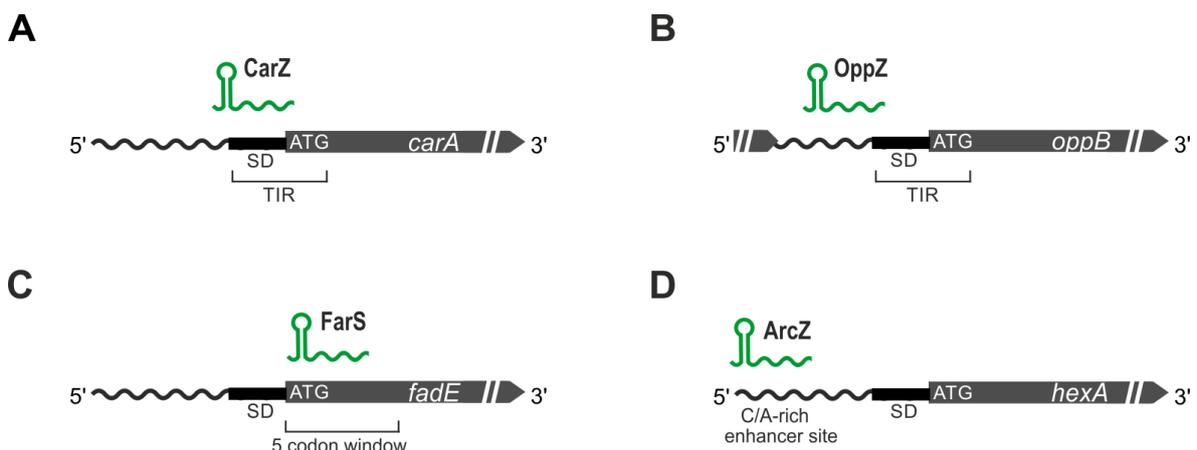
## 6.2 Target regulation by Hfq-binding sRNAs

Hfq is the most extensively studied bacterial RNA chaperone and affects gene expression globally by stabilizing sRNAs and promoting their interactions with target mRNAs. The stability of numerous sRNAs has been shown to be strongly reduced in the absence of Hfq (30, 39, 62, 115), and for many sRNAs Hfq is even required for biogenesis (35). In this thesis, several Hfq-binding sRNAs have been studied and we investigated their Hfq dependence in detail. Regarding the role Hfq here plays for stability and processing, there are differences between the studied sRNAs: In our experiments, FarS and OppZ were still detectable in an *hfq* mutant (chapter 2, Fig. S5A; chapter 3, Fig. 2 – suppl. 2), whereas CarZ and ArcZ were not (data not shown for CarZ; chapter 4, Ext. Data Fig. 1). This suggests for FarS and OppZ that, although both are produced from longer transcripts by RNase E - mediated cleavage, Hfq is not required for initial maturation. After initial biogenesis of FarS and OppZ, the stability of both sRNAs, however, is strongly reduced in the absence of Hfq (chapter 2, Fig. S5A; chapter 3, Fig. 2 – suppl. 2). In contrast, for CarZ it might be that Hfq is required to guide RNase E for initial processing from the parental mRNA 3' UTR. For ArcZ, RNase E-mediated processing of the full length transcript into a shorter stable form has been examined in detail by *in vitro* experiments in *Salmonella* (35).

In addition to stabilizing sRNAs, Hfq acts as a “matchmaker” between sRNAs and their targets, and thereby enables them to act as post-transcriptional regulators. Although the principle of short and imperfect pairing is a general feature of Hfq-binding sRNAs (1), there are multiple possibilities for the exact regulatory mechanism. The sRNAs studied in chapter 2 – 4 of this thesis (FarS, OppZ, CarZ and ArcZ) all post-transcriptionally repress their respective mRNA targets in an Hfq-dependent manner (chapter 2, Fig. S5 B and C; chapter 3, Fig. 3 – suppl. 2 A and B, Fig. 7 – suppl. 1 D and E), but achieve this through binding at different positions of the mRNAs. CarZ is an example for a “prototypic” sRNA directly sequestering the Shine-Dalgarno sequence (Fig. 6.4 A) and thereby most likely inhibiting the first step of translation initiation, the binding of the 30S ribosomal subunit. The OppZ base-pairing region is further upstream of the *oppB* start codon (-30 to -9) (Fig. 6.4 B), but still within a distance where direct inhibition of translation initiation is very likely, given that

the ribosome occupies ~20 nucleotides upstream of the start codon (214, 215). Translation inhibition, however, can also be achieved by sRNAs pairing downstream of the start codon, as it has been shown first for the RybB-*ompN* sRNA-target pair in *Salmonella* (216). The FarS sRNA studied here indeed binds on both *fadE* mRNAs within the coding sequence (chapter 2, Fig. S5 D and E). The validated regions of interaction (chapter 2, Fig. 4B-E) are located in the previously predicted “5 codon window” downstream of the start codon (Fig. 6.4 C), where translation initiation might be inhibited by steric hindrance of the ribosome (216).

Although these scenarios where sRNAs operate closely to the translation initiation region (TIR) seem to be the case for the majority of sRNA-target pairs, recently, more and more variations have been discovered, where sRNAs repress their targets from a larger distance (reviewed in (41)). For instance, the Spot 42 sRNA base-pairs with the *sdhC* mRNA too far upstream of the TIR to directly inhibit ribosome association. Instead, inhibition of translation initiation is achieved by guiding Hfq close to the SD sequence where the protein inhibits ribosome binding (217). Moreover, sRNAs can act as repressors by binding on ribosome standby sites (44). These are sites of nonspecific ribosome binding facilitating access of the ribosome when structured RBS transiently open (218, 219). Another possibility is base-pairing to so called translational enhancer elements. This has been reported for example for the GcvB sRNA, targeting the *gltI* mRNA in a conserved C/A-rich region about 50 nucleotides upstream of the start codon (42). The site of base-pairing between ArcZ and the *hexA* mRNA identified here (chapter 4, Fig. 3A and Suppl. Fig. 5) is about the same distance, likewise C/A-rich and equally well conserved, suggesting that ArcZ functions by such a mechanism (Fig. 6.4 D).



**Figure 6.4: Repression of translation initiation by base-pairing sRNAs**

**A)** CarZ base-pairs to the SD sequence of *carA*. **B)** OppZ base-pairs in the IGR between *oppA* and *oppB*. **C)** FarS base-pairs in the CDS of the *fadE* mRNAs. **D)** ArcZ base-pairs to a C/A-rich enhancer region in the 5' UTR of *hexA*.

The consequence of all the described base-pairings here is downregulation of target gene expression. However, the fate of the target as well as the regulator RNA molecule upon the interaction can differ, which strongly influences the regulatory dynamics of sRNA-target circuits. In the absence of translation, mRNAs are exposed to ribonucleolytic attack and are usually rapidly degraded. A common scenario is that sRNA and mRNA are both degraded upon the interaction (coupled degradation) (115, 220). However, sRNAs can also be recycled while only the target mRNA is degraded (catalytic degradation) (88) or the two molecules sequester each other without any degradation (87, 221). The outcome of a base-pairing interaction has tremendous consequences for further regulation, as the ratio between sRNA and target molecules is a crucial factor. A possibility to test the effect of the binding interaction between two RNA molecules on their degradation is to analyze stability vice versa upon overexpression of the other factor (87, 88). For *oppB* and *carAB* mRNAs, we tested stability upon short overexpression of the OppZ and CarZ sRNAs, respectively. Interestingly, the stability of *carAB* (chapter 3, Fig. 7 – suppl. 2 A and B) here is clearly reduced, whereas stability of *oppB* is not significantly altered (chapter 3, Fig. 3 – suppl. 1D). The reason for this could be that in the case of OppZ, binding in the intergenic region between *oppA* and *oppB*, translation of the upstream OppA stabilizes the transcript. Our results further showed that translation inhibition of *oppB* leads to another secondary regulatory effect, namely premature transcription termination (chapter 3, Fig. 5), and our data point to the same mechanism for *carAB*. It might be possible that the reduction of stability of *carAB* in contrast to *oppB* additionally contributes to the stronger downregulation observed in the translational reporter fusion of *carA* in comparison to *oppB* (chapter 3, Figs. 7C and 3D).

All the scenarios described so far represent a base-pairing interaction between an sRNA and an mRNA, resulting in altered protein levels of the target. However, regulation by Hfq-binding sRNAs can also involve duplex formation between two non-coding RNA molecules. As discussed in section 6.1.2, also due to novel methodologies, recently several such interactions mediated by Hfq have been identified or predicted (126, 171). The sRNA studied in chapter 5 (QrrS) is engaged in this type of interaction, and in this regard differs from the other sRNAs studied in this thesis, as it base-pairs with other sRNAs but not with mRNA targets. With these interactions, sRNAs can regulate the levels or the activity of other sRNAs, consequently affecting expression of mRNA targets of the latter (67, 209). Also in this context, it is a fundamental difference, whether the interaction yields in degradation of one or both binding partners. For the interaction identified in chapter 5, the base-pairing between QrrS and the homologous sRNAs Qrr1-4, our results suggest that the regulation is directional: Stability of Qrr4 is strongly reduced upon pairing of QrrS, but not vice versa (chapter 5, Fig. 3). This has important consequences for the regulatory dynamics of the *V. cholerae* QS system (see Fig. 1.2), as Qrr1-4 molecules consequently are not available

for regulating their mRNA targets. The setting here is in particular intriguing, because the Qrr1-4 sRNAs also undergo different fates depending on which of the mRNA targets they are binding to (87). Hence, the question arises of what exactly determines the outcome after duplex formation. *Feng et al.* (87) found that the strength of the interaction here clearly is at least one crucial factor, but further investigation would be needed to overall understand the underlying molecular principles here.

### 6.3 Origin and evolution of sRNAs and the role of Hfq and RNase E

The structural and functional flexibility of sRNAs and the only limited complementarity to target sequences required to carry out regulatory functions allow a rapid evolution of sRNAs. Generally, different scenarios have been reported to be the source of new sRNA genes (reviewed in (222)). For instance, several sRNAs have been acquired through horizontal gene transfer from bacteriophages or plasmids. Gene duplications or genomic rearrangements can be further origins for sRNA evolution (222). In principle, a single point mutation may be sufficient to establish a base-pairing interaction with a target RNA and thereby provide a selective advantage for a transcript to be maintained in evolution. Target regulation by the *Salmonella* SgrS sRNA shows impressively how one nucleotide can make the difference: SgrS represses translation of the *sopD* mRNA, whereas *sopD2*, which evolved after gene duplication, is not regulated by the sRNA due to a single nucleotide difference in the relevant region for base-pairing (223).

We have learned that sRNAs can origin from almost all segments of the genome, and an increasing number of sRNAs produced from UTRs have been discovered (30, 34). These are in particular interesting from an evolutionary perspective. Depending on the genomic location, different evolutionary mechanisms may play a role. For sRNAs expressed from 5' UTRs, the scenario could be that the sequence which now constitutes an sRNA was once the 5' UTR of a coding sequence, and over time mutations generated a stable terminator (222). The fact that riboswitches exist which function as *trans*-acting sRNAs support this hypothesis (224). However, more than 5' UTRs, 3' UTRs seem to be ideal loci for the evolution of new sRNAs. Here, the presence of a Rho-independent terminator, one of the essential features of Hfq-binding sRNAs, might play the crucial role. The affinity of Hfq towards Rho-independent terminators seems to be a driving force for the evolution of sRNAs from 3' UTRs and an important factor for the apparently high number of 3' UTR derived sRNAs (34, 222). Among the sRNAs detected to bind to Hfq in our RIP-seq experiments in

chapter 2 indeed more than 50% are located in the 3' UTR of coding sequences (chapter 2, Fig. 2A). Similarly, in *Salmonella* Hfq co-immunoprecipitation experiments, many 3' UTRs with predicted Rho-independent terminators were significantly enriched, suggesting a large number of possible candidates for novel 3' UTR derived sRNAs (30). In evolutionary terms, mutations here could lead to a promoter driving expression of the sRNA independently from the parental gene, or could generate RNase E cleavage sites, yielding an sRNA which is clipped-off from the mRNA.

Indeed, accumulating evidence exists that RNase E, in conjunction with Hfq, is a main player for the biogenesis of many sRNAs in Gram-negative bacteria (35), and our results in chapter 3 confirm that this is also the case in *V. cholerae*. Determining RNase E cleavage sites genome-wide helps to identify sRNAs generated by processing from mRNA transcripts. Having this information gives important hints regarding the regulatory function of the sRNA, as it implicates that expression of the sRNA depends on the transcriptional control of the parental mRNA rather than being independently regulated by an internal promoter. For *V. cholerae*, we found a set of 17 3' UTR derived sRNAs (from previously annotated candidates (39)) with potential RNase E cleavage sites at the predicted starts of the sRNAs (chapter 3, Fig. 1 – suppl. 4 and Supplementary file 2A), suggesting an RNase E - dependent maturation. Additionally, for the newly identified sRNAs in chapter 2, the TIER-seq data predict RNase E cleavage sites at the putative 5' end of six more sRNAs (Vcr203, Vcr213, Vcr216, Vcr219, Vcr222, Vcr228). The actual number of 3' UTR derived sRNAs generated by RNase E - dependent processing might be even higher, as very stable sRNAs could have escaped detection due to the only transient inactivation of RNase E in these experiments, and moreover, the data can only represent a snapshot at a given condition (35).

Overall, it has become ever more clear that the functioning and evolution of sRNAs is closely linked to the presence of Hfq and RNase E, raising the question of how conserved the functions of these proteins are and how the characteristics and roles of sRNAs differ in bacteria without Hfq or RNase E. While RNase E is known as a major endoribonuclease in Gram-negative bacteria, its role in Gram-positive bacteria is not well defined (225). Hfq as well clearly does not have the same role in Gram-positive bacteria. On the contrary, although there are Hfq homologs in about half of all sequenced species (90, 91), the function of Hfq so far remains rather unclear in Gram-positive bacteria and only few sRNAs seem to be Hfq-dependent in these organisms (226). The prime example here is the sRNA LhrA in *Listeria monocytogenes*, which has been demonstrated to require Hfq for target regulation (227). Given this completely different picture in Gram-positive bacteria, it might be interesting to study from a more global perspective the features and genomic locations of sRNAs in these organisms. For *Bacillus subtilis*, for example, regulatory RNA elements have been predicted and categorized genome-wide, showing a large variety of possible RNA-based regulatory

mechanisms (228, 229). Furthermore, pervasive antisense transcription has been observed here as well as in studies on other Gram-positive bacteria, such as *S. aureus* and *L. monocytogenes* (230, 231). Related to this, intriguingly, also long and conserved 3' UTRs have been reported (226, 230-232). It could be that these are involved in regulatory functions, possibly in conjunction with other undiscovered RNA-binding proteins, which potentially but not necessarily might have similar functions as Hfq in Gram-negative bacteria. Given the apparently less important role of Hfq in Gram-positive bacteria, it is also intriguing to ask to what extent Hfq homologs differ. Remarkably, there are indications for some connection between the presence or absence of conserved arginines on the rim of Hfq, which have been shown in *E. coli* to play an important role in mediating base-pairing interactions, and the relevance Hfq plays as a chaperone in various Gram-positive species, including *L. monocytogenes*, *B. subtilis* and *S. aureus* (109). Altogether, it is clear that insights into sRNA-based regulation derived from studies in Gram-negative model organisms, where Hfq is a global actor, cannot simply be transferred to Gram-positive bacteria.

### 6.4 ArcZ is a highly conserved sRNA

The rapid evolution of sRNAs makes it even more interesting to look at their distribution and conservation in diverse species. Whereas some sRNAs show quite a limited distribution, others are broadly conserved (222). Intriguingly, many of the characterized sRNAs appeared after the separation of the *Enterobacteriales* from the other orders of the gamma-proteobacteria (233). There are only few sRNAs known which are conserved outside the *Enterobacteriales* (233). An example for this is the Spot 42 sRNA, regulating carbohydrate metabolism, which has also been found in four other orders of gamma-proteobacteria, including the *Vibrionales* (234).

One highly conserved sRNA in the *Enterobacteriales* order is the ArcZ sRNA, studied in chapter 4. Therefore, ArcZ represents an ideal example to have a closer look at conservation of sRNAs and their interactions with targets. Together with other conserved sRNAs involved in the regulation of central cellular functions, e.g. RybB, ArcZ is often described as one of the enterobacterial “core” sRNAs (37). Alignment of *arcZ* sequences from various species showed a high degree of conservation, with the 3' region being the most conserved part (chapter 4, Fig. 1A, and (37)). Similar to *Salmonella* and *E. coli*, where ArcZ has been studied most extensively (37, 65), in *P. laumondii* and *X. szentirmaii*, too, ArcZ is transcribed as a ~120 nt long sRNA and processed into a ~50 nt long stable shorter

transcript (chapter 4, Figs. 2A and 3D). The genomic context of the ArcZ sRNA is equally well conserved, it is consistently located between the *yhbL* and *arcB* genes or their respective homologs (37).

Regarding the expression of the sRNA, ArcZ is expressed at all stages of growth in *P. laumondii* and is most abundant at high cell density (chapter 4, Fig. 2A). A similar pattern has been observed in *Salmonella* (37). However, strikingly, in *X. szentirmaii*, which is very closely related to *P. laumondii*, we detected a decrease in expression towards the 24 h time-point (chapter 4, Fig. 3D). This was not the case in *P. laumondii*, indicating some different regulatory dynamics. The phenomenon of altered regulation of conserved sRNAs has also been observed previously for other sRNAs (222). Interestingly, even between the closely related species *Salmonella* and *E. coli*, there seem to be minor differences in the expression of ArcZ (previously identified as SraH) regarding abundance at low cell density (24, 37, 65). Overall, although ArcZ is a very conserved sRNA, the different expression patterns here might reflect fine-tuning of gene regulation and adaptation to specific environmental conditions.

As the regulatory function of an sRNA depends on its interaction with a target RNA, it is obvious to study conservation of sRNAs in conjunction with conserved base-pairing regions of putative targets. Our data show that ArcZ base-pairs to and represses the *hexA* mRNA in *P. laumondii* (chapter 4, Fig. 3 A and B). The region of interaction, which we validated by compensatory base pair exchange (chapter 4, Fig. 3 B and C), is mostly identical to the region of ArcZ involved in base-pairing with targets in *Salmonella* and *E. coli* (37, 65). Three targets that have been identified and for which direct pairing with ArcZ has been confirmed in *Salmonella* are *sdaC*, *tpx* and STM3216, encoding a serine transport protein, a thiol peroxidase and a methyl-accepting chemotaxis protein, respectively (37). Although homologs of these exist in *P. laumondii* (*plu1578*, *plu2579* and *plu1853*), according to our data, they do not seem to be repressed by ArcZ in this species under the tested condition (chapter 4, Suppl. Table 8). In fact, when looking at the proposed base-pairing interactions in *Salmonella* (37), the involved mRNA regions are not conserved in *P. laumondii*. This might be an example where the function of the sRNA diverged. A similar situation can be observed when examining targets of the conserved RyhB and RybB sRNAs: Only for a minor fraction of targets found in *E. coli* the interaction is also conserved in *P. laumondii* (235). In this context, it is also noteworthy that *rpoS*, encoding a major enterobacterial stress sigma factor, is activated by ArcZ in *E. coli*. Intriguingly, there is a single nucleotide difference between *E. coli* and *P. laumondii* in the region of the *rpoS* 5' UTR which is involved in duplex formation with ArcZ in *E. coli*. It would therefore be interesting to experimentally test whether or not there is a direct interaction between ArcZ and *rpoS* in *P. laumondii*. Altogether, remarkably, against what one might expect, *Richter and Backofen*

(235) revealed that, although the base-pairing regions of sRNAs are highly conserved, the regions of interactions in mRNA targets do not show high sequence conservation. Conservation of target regulation across related species does hence not seem to be a general rule (235).

Nevertheless, the region of the *hexA* mRNA base-pairing to ArcZ in *P. laumondii* is very conserved in related bacteria, including *Serratia* and *Erwinia* species (chapter 4, Suppl. Fig. 5). This suggests that here as well ArcZ might bind to *hexA*. As these bacteria also produce secondary metabolites (236, 237), it is tentative to speculate that ArcZ, post-transcriptionally repressing the mRNA for the master regulator HexA, is a key factor for secondary metabolite production across several species. In accordance with this, in *Erwinia*, HexA has previously been shown to negatively regulate the production of exoenzymes, secondary metabolites and signals for cell-to-cell communication, and to be involved in virulence (238, 239). Moreover, it has been demonstrated that Hfq is key for virulence in the plant pathogen *Erwinia amylovora* and that indeed ArcZ is involved here, but direct targets of ArcZ had not been identified in this study (240). Similarly, in *Serratia*, a genus belonging to the *Yersiniaceae*, Hfq has been revealed as a crucial factor for the production of several natural products, but the exact mechanisms for the loss of secondary metabolite production in an *hfq* mutant remained elusive (241, 242). Altogether, although further investigation would be needed to determine in detail the regulatory mechanisms in these related bacteria, with the identification of the direct interaction between the Hfq-binding sRNA ArcZ and the *hexA* mRNA in *P. laumondii*, our study unraveled an important link and might help to clarify previously observed phenotypes.

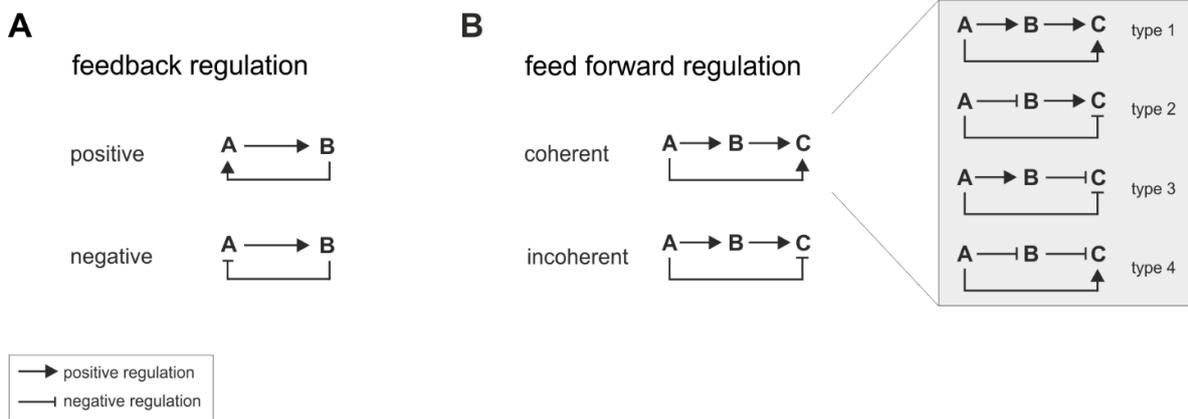
## 6.5 Integration of sRNAs in regulatory circuits

The evolutionary pressure to fine-tune regulation of gene expression and the versatility of sRNAs makes them ideal factors to be part of regulatory networks, expanding the possibilities of complex regulation. Typically, sRNAs are expressed in response to specific environmental signals, such as nutrient limitation, temperature or other stress factors, and are frequently involved in circuits together with transcription factors (75). Here, sRNAs can act at different levels of a regulatory cascade and can be involved in various regulatory motifs, often resulting in tight and nuanced feedback or feed forward regulation (73, 74).

### 6.5.1 Autoregulatory feedback regulation at the RNA level

In feedback loops, a regulator controls its own expression. Feedback loops can be either positive or negative (Fig. 6.5 A), depending on whether the regulator enhances or represses its expression (243). For transcriptional regulators, in particular the dynamics of autoregulatory feedback loops have been studied extensively (243, 244). However, it is nowadays clear that sRNAs are also important factors for feedback regulation, forming mixed regulatory circuits together with transcription factors (73, 74). An example for mixed feedback regulation involving several factors is the connection between the transcriptional regulator Fur, the RyhB sRNA and multiple iron containing proteins. RyhB is repressed by Fur when the intracellular concentration of iron is high. When iron is scarce, RyhB post-transcriptionally downregulates several mRNAs encoding iron-utilizing proteins. This increases the intracellular iron concentration, resulting again in repression of RyhB by Fur (56, 57).

Several other sRNAs have been reported to be part of feedback loops (e.g. OmrA/B in *E. coli*, Qrr1-4 in *V. cholerae*), however, these always involved a transcription factor (84, 245). It was assumed that autoregulatory feedback regulation is a mechanism only possible for transcription factors and that sRNAs cannot regulate their own expression directly (73). However, our study in chapter 3 demonstrates that sRNAs processed from 3' UTRs of mRNAs are indeed able to regulate their own expression. Binding of OppZ to the IGR between *oppA* and *oppB* followed by translation repression results in premature transcription termination of the *oppABCDF* operon through Rho. As OppZ is produced by RNase E - mediated cleavage from the parental mRNA, the transcription termination within the *oppB* gene results in decreased OppZ sRNA levels (chapter 3, Fig. 5). Moreover, we find a second sRNA (CarZ) base-pairing with and repressing the mRNA from which it is processed (chapter 3, Fig. 7), suggesting that this could be a more widespread mechanism. Although understanding the exact biological functions of both autoregulatory circuits would require further investigation, the novel concept identified here shows how sRNAs can in fact regulate their own expression without any transcription factor being involved. Similar to what has been described for feedback autoregulation at the transcriptional level (244), our results indicate that one benefit here might be accelerated response time (chapter 3, Fig. 8). In case of the OppZ sRNA, the mechanism additionally allows regulation at the suboperonic level, with expression of the upstream *oppA* being unaffected (chapter 3, Fig. 5), which could not be achieved by sole transcriptional regulation.



**Figure 6.5: Regulatory network motifs**

**A)** Positive and negative feedback regulation involving two factors. **B)** Coherent and incoherent feed forward regulation. For coherent regulation, all possible combinations of activating and repressing factors are visualized (gray box); for incoherent regulation, only type 1 is shown for the sake of simplicity.

### 6.5.2 A feed forward loop allowing nuanced regulation of *V. cholerae*'s fatty acid metabolism

In feed forward loops (FFLs), expression of a target is controlled both directly by a first regulator (transcription factor or sRNA) and indirectly via a second regulator, which itself also is controlled by the first one. Two different types of feed forward regulation can be distinguished: coherent and incoherent. FFLs are called coherent, when both regulatory arms are either activating or repressing the target; and are called incoherent, when one is activating and the other is repressing (246). Following *Mangan and Alon* (247), FFLs can be further classified according to the different combinations of activating and repressing factors, resulting in total in eight distinct subtypes (four coherent and four incoherent) (Fig. 6.5 B), implying different dynamic properties.

Similar to feedback regulation, benefits and dynamics of feed forward loops have been mostly studied at the transcriptional level (247, 248). Nevertheless, several mixed FFLs including sRNAs at either the top or the middle position have been reported (reviewed in (74)). For coherent FFLs, which are the focus here, examples including an sRNA have indeed been found for all four subtypes (Fig. 6.5 B): The regulation of *flhD*, encoding a master regulator for flagella synthesis in *E. coli*, by the global transcription factor Crp and the sRNA McaS is an example for type 1, meaning that Crp activates McaS, and both Crp and McaS activate *flhD*, at the transcriptional and at the post-transcriptional level, respectively (249). An example where the top regulator represses the target as well as an

activating regulator in the middle (type 2) is the RNAIII-Rot-*spa* loop in *S. aureus*. Here, the sRNA, RNAIII, acts on the top position, downregulating the target mRNA *spa* at high cell density as well as its transcriptional activator Rot, preventing transcriptional leakage and ensuring fine-tuned regulation (250, 251). In the osmotic stress response in *E. coli*, the transcription factor OmpR is even involved in both type 3 and type 4 FFLs involving an sRNA in the middle (type 3: OmpR-MicF-*ompF* (23, 252, 253); type 4: OmpR-MicC-*ompC* (253, 254)).

The FadR-FarS-*fadE* feed forward loop identified in this thesis (chapter 2), according to this classification, constitutes a mixed coherent type 3 loop. It is novel in the sense that it involves an sRNA which is processed from the 3' UTR of an mRNA. The transcription factor FadR here acts on top and the sRNA FarS acts as regulator in the middle (chapter 2, Fig. 6A). FadR activates expression of FarS via transcriptional activation of *fabB*, and both factors, FadR and FarS, repress the paralogous *fadE* mRNAs. FadR is a conserved transcription factor in enteric bacteria belonging to the GntR family and binding to genes of the fatty acid biosynthesis and degradation pathways (255). In the absence of exogenous long-chain fatty acids (LCFAs), FadR binds to the promoter of the *fabB* gene (256, 257), activating expression of *fabB*-FarS (chapter 2, Fig. 2 D and E, Fig. S3B). In addition, FadR inhibits transcription of several genes of the fatty acid degradation pathway, including the *fadE* genes (256, 258). When LCFAs are imported into the cell, the subsequent binding to FadR results in a conformational change of the transcription factor abolishing its binding to the respective promoters (259).

Major benefits of mixed feed forward loops that have been described include stricter, double-layered regulation and enhanced response (74, 253). Double-layered regulation means that control of gene expression occurs at the transcriptional as well as at the post-transcriptional level, which can provide an overall tighter control, as for example shown for the above mentioned regulation of *ompF* in the OmpR-MicF-*ompF* FFL (253). With FarS, the *fadE* genes are not only regulated by the transcription factor FadR, but also at the level of translational repression. The *fadE* genes, encoding coA-dehydrogenase, play important roles in the degradation pathway of fatty acids in *V. cholerae*. The enzyme catalyzes the first step in the  $\beta$ -oxidation cycle after exogenous fatty acids are imported into the cell. Our data show that FadE protein levels are indeed elevated in a  $\Delta farS$  mutant strain (chapter 2, Fig. 5 A and B, Fig. S6 A and B), and moreover, we observed altered response dynamics under changing fatty acid concentrations. When cells are exposed to a sudden increase in exogenous fatty acids, our results indicate that FarS serves as a delay element (chapter 2, Fig. 6 A and B, Fig. S6 E and F), a characteristic property of coherent FFLs (74, 247). In the reverse scenario, when cells are exposed to a sudden decrease in fatty acids, FarS speeds up the repression of FadE (chapter 2, Fig. 6 C and D, Fig. S6 G and H). Dynamics of mixed

## 6 DISCUSSION

FFLs can be further analyzed by mathematical modeling (253), which could also be interesting for the FadR-FarS-*fadE* loop, but this would first require the determination of additional parameters, such as the affinity of FadR to the *fabB* and *fadE* promoters. Generally, tight control of fatty acid metabolism is crucial for bacteria as fatty acids are important components for membrane lipid homeostasis as well as for energy metabolism (260). Therefore, biosynthesis and degradation pathways must be regulated dynamically in response to the availability of fatty acids in the surrounding, and FarS turned out to be a central player for this in *V. cholerae*.

Given the relatively conserved function of FadR in gamma-proteobacteria (261), the question arises whether FarS is also conserved in species outside the *Vibrionaceae*, for example in *E. coli*, or if there is another sRNA with a similar function. Analysis of the sequence of the *fabB* 3' UTR in *E. coli* did not reveal a FarS homolog neither is there an obvious complementary sequence prone for base-pairing between the *fabB* 3' UTR and the *fadE* 5' UTR. However, it is noteworthy that the *fabB* 3' UTR seems to be a strong binding partner of ProQ in *E. coli* and *Salmonella* (136), suggesting that regulatory functions of the *fabB* 3' UTR might have evolved independently here. Related to this, another difference between *E. coli* and *V. cholerae* is that *E. coli* has only one *fadE* gene encoding acyl-coA-dehydrogenase, whereas the gene is duplicated in *V. cholerae*; and given that the FarS binding site is highly conserved in both paralogs (chapter 2, Fig. S5 D and E), one could speculate that the duplication occurred after the regulation via FarS was established. Further, there is an important difference between the *E. coli* and the *V. cholerae* FadR protein. The FadR protein of the *Vibrionaceae* is unusually long and has a second binding site for long-chain acyl-CoAs, allowing a tighter regulation (256, 261). Overall, this suggests that here different mechanisms have evolved to achieve an elaborate regulation of fatty acid metabolism and adaptation to specific environmental conditions.

Regarding the natural environment of *V. cholerae*, fatty acid concentrations can change dramatically during the bacterium's lifecycle, switching between aquatic surroundings and the human host. It has been shown that *Vibrios* can obtain fatty acids from the sediment in aquatic ecosystems as well as the uptake of host-derived fatty acids plays an important role during colonization (262, 263). Research by Giles *et al.* (263) demonstrated that polyunsaturated fatty acids, which are components of bile, present in the small intestine, where *V. cholerae* colonizes, are incorporated into phospholipids. The study suggests that *Vibrios* can take up a broader range of exogenous fatty acids than other bacteria. A further link between the uptake of fatty acids and pathogenesis is provided by the finding that genes of the degradation pathway including *fadL*, encoding the fatty acid transporter in the outer membrane, and the two *fadE* paralogs were found to be upregulated during infection (264). Interestingly, it was discovered very recently that also expression of the gene upstream of

*vc1740* (*fadE*), with which it forms a di-cistronic operon, is increased during infection (265). This gene, *vc1741*, which we showed is not affected by the post-transcriptional repression via FarS (chapter 2, Fig. 4D), encodes a transcription factor also binding to the promoters of several genes of the *fad* regulon and responding to LCFAs (265). These findings add new details to the sophisticated control of fatty acid metabolism in *V. cholerae*. In summary, tight and dynamic regulation here seems to be crucial for *V. cholerae*'s lifestyle and pathogenicity, and the feed forward loop involving the FarS sRNA identified in this thesis contributes to this.

## 6.6 Summary and perspective

The aim of the study was to identify and characterize small regulatory RNAs associating with the RNA chaperone Hfq, both on a global scale as well as in detail for selected examples. The focus here was on the model organism and human pathogen *V. cholerae*. The results presented in this thesis significantly add to our knowledge on Hfq-binding sRNAs in this organism. The global studies on association of RNAs with Hfq (RIP-seq and RIL-seq) provide comprehensive data sets which can serve as valuable resources for the whole research community for further investigation on regulatory RNAs in *V. cholerae*. Moreover, the data sets are not only useful for the *Vibrio* community; having transcriptome-wide information about Hfq-ligands in various organisms enhances the overall understanding of the impact of Hfq-mediated post-transcriptional regulation. In particular, the relatively novel methodology of RIL-seq has great potential for the discovery of new aspects of sRNA-based regulation of gene expression. For instance, it might help to foster our understanding of the role RNA-RNA interactions play which do not result in up- or downregulation of target genes.

Apart from these global approaches, several *V. cholerae* sRNAs have been investigated in great detail and thereby novel aspects regarding functional as well as physiological roles have been revealed. We discovered that the FarS sRNA is a central part of an RNA-based feed forward loop which for the first time involves an sRNA derived from the 3' UTR of a coding sequence, and it is one of the very few reported bacterial sRNAs involved in central fatty acid metabolism. The study on OppZ and CarZ demonstrates for the first time how autoregulatory feedback regulation is possible at the RNA level without involving any transcription factor. We expect that this is a more widespread phenomenon, opening new exciting questions about the dynamics of autoregulation, an ubiquitous regulatory principle in biology. The QrrS sRNA discovered here is very particular in the sense that it base-pairs with four homologous sRNAs, but it is even more astounding regarding its

## 6 DISCUSSION

physiological role in the quorum sensing system of *V. cholerae*. Although *Vibrios* have served as model organisms for the investigation of bacterial communication, this sRNA sponge had not been found previously and sheds new light on the dynamics of this central aspect of *V. cholerae*'s lifestyle.

The second organism studied in this thesis with respect to Hfq-binding sRNAs was *P. laumondii*. Here, the RIP-seq data set might be in particular valuable for other researchers as very little is known so far about Hfq-binding sRNAs in this organism. The study further is novel in addressing sRNAs also in the closely related bacterium *X. szentirmaii*. Moreover, the discovery that the Hfq-binding sRNA ArcZ binds to *hexA* and thereby regulates specialized metabolite production in *P. laumondii*, and possibly in other secondary metabolite producing bacteria too, is a key finding of this study and might open new pathways in the field of natural product biosynthesis.

In conclusion, the findings of this thesis contribute significantly to our understanding of the functioning of bacterial sRNAs and provide useful resources for future studies in this booming research field, which in the last two decades uncovered the exciting importance of RNA-based gene regulation. Detailed understanding of the mechanistic principles of sRNAs not only helps to uncover different aspects of bacterial physiology, but might also allow to use these versatile molecules for different RNA-based applications, including synthetic and therapeutic purposes.

# References

(for chapters 1 and 6)

1. Storz G, Vogel J, & Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell* 43(6):880-891.
2. Simons RW & Kleckner N (1988) Biological regulation by antisense RNA in prokaryotes. *Annu Rev Genet* 22:567-600.
3. Hor J, Gorski SA, & Vogel J (2018) Bacterial RNA Biology on a Genome Scale. *Mol Cell* 70(5):785-799.
4. Morris KV & Mattick JS (2014) The rise of regulatory RNA. *Nat Rev Genet* 15(6):423-437.
5. Richards GR & Vanderpool CK (2011) Molecular call and response: the physiology of bacterial small RNAs. *Biochim Biophys Acta* 1809(10):525-531.
6. Kortmann J & Narberhaus F (2012) Bacterial RNA thermometers: molecular zippers and switches. *Nat Rev Microbiol* 10(4):255-265.
7. Loh E, Righetti F, Eichner H, Twittenhoff C, & Narberhaus F (2018) RNA Thermometers in Bacterial Pathogens. *Microbiol Spectr* 6(2).
8. Grosso-Becera MV, Servin-Gonzalez L, & Soberon-Chavez G (2015) RNA structures are involved in the thermoregulation of bacterial virulence-associated traits. *Trends Microbiol* 23(8):509-518.
9. DiRita VJ, Parsot C, Jander G, & Mekalanos JJ (1991) Regulatory cascade controls virulence in *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 88(12):5403-5407.
10. Champion GA, Neely MN, Brennan MA, & DiRita VJ (1997) A branch in the ToxR regulatory cascade of *Vibrio cholerae* revealed by characterization of *toxT* mutant strains. *Mol Microbiol* 23(2):323-331.
11. Weber GG, Kortmann J, Narberhaus F, & Klose KE (2014) RNA thermometer controls temperature-dependent virulence factor expression in *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 111(39):14241-14246.
12. Lotz TS & Suess B (2018) Small-Molecule-Binding Riboswitches. *Microbiol Spectr* 6(4).
13. McCown PJ, Corbino KA, Stav S, Sherlock ME, & Breaker RR (2017) Riboswitch diversity and distribution. *RNA* 23(7):995-1011.
14. Thomason MK & Storz G (2010) Bacterial antisense RNAs: how many are there, and what are they doing? *Annu Rev Genet* 44:167-188.
15. Brantl S (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* 10(2):102-109.
16. Masachis S & Darfeuille F (2018) Type I Toxin-Antitoxin Systems: Regulating Toxin Expression via Shine-Dalgarno Sequence Sequestration and Small RNA Binding. *Microbiol Spectr* 6(4).
17. Waters LS & Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136(4):615-628.
18. Romeo T & Babitzke P (2018) Global Regulation by CsrA and Its RNA Antagonists. *Microbiol Spectr* 6(2).
19. Liu MY, *et al.* (1997) The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem* 272(28):17502-17510.
20. Weilbacher T, *et al.* (2003) A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol Microbiol* 48(3):657-670.
21. Kawamoto H, Koide Y, Morita T, & Aiba H (2006) Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol* 61(4):1013-1022.

## REFERENCES

22. Gorski SA, Vogel J, & Doudna JA (2017) RNA-based recognition and targeting: sowing the seeds of specificity. *Nat Rev Mol Cell Biol* 18(4):215-228.
23. Mizuno T, Chou MY, & Inouye M (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc Natl Acad Sci U S A* 81(7):1966-1970.
24. Argaman L, *et al.* (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 11(12):941-950.
25. Wassarman KM, Repoila F, Rosenow C, Storz G, & Gottesman S (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* 15(13):1637-1651.
26. Tjaden B, *et al.* (2002) Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res* 30(17):3732-3738.
27. Chen S, *et al.* (2002) A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems* 65(2-3):157-177.
28. Adams PP & Storz G (2020) Prevalence of small base-pairing RNAs derived from diverse genomic loci. *Biochim Biophys Acta Gene Regul Mech*:194524.
29. Sittka A, *et al.* (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* 4(8):e1000163.
30. Chao Y, Papenfort K, Reinhardt R, Sharma CM, & Vogel J (2012) An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* 31(20):4005-4019.
31. Heidrich N, *et al.* (2017) The primary transcriptome of *Neisseria meningitidis* and its interaction with the RNA chaperone Hfq. *Nucleic Acids Res* 45(10):6147-6167.
32. Sharma CM, *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464(7286):250-255.
33. Kroger C, *et al.* (2012) The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci U S A* 109(20):E1277-1286.
34. Miyakoshi M, Chao Y, & Vogel J (2015) Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Curr Opin Microbiol* 24:132-139.
35. Chao Y, *et al.* (2017) In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Mol Cell* 65(1):39-51.
36. Chao Y & Vogel J (2016) A 3' UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. *Mol Cell* 61(3):352-363.
37. Papenfort K, *et al.* (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol Microbiol* 74(1):139-158.
38. Altuvia S, Zhang A, Argaman L, Tiwari A, & Storz G (1998) The *Escherichia coli* OxyS regulatory RNA represses *fhfA* translation by blocking ribosome binding. *EMBO J* 17(20):6069-6075.
39. Papenfort K, Forstner KU, Cong JP, Sharma CM, & Bassler BL (2015) Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proc Natl Acad Sci U S A* 112(7):E766-775.
40. Romilly C, *et al.* (2014) A non-coding RNA promotes bacterial persistence and decreases virulence by regulating a regulator in *Staphylococcus aureus*. *PLoS Pathog* 10(3):e1003979.
41. Frohlich KS & Papenfort K (2020) Regulation outside the box: new mechanisms for small RNAs. *Mol Microbiol*.
42. Sharma CM, Darfeuille F, Plantinga TH, & Vogel J (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* 21(21):2804-2817.
43. Azam MS & Vanderpool CK (2020) Translation inhibition from a distance: The small RNA SgrS silences a ribosomal protein S1-dependent enhancer. *Mol Microbiol*.

44. Darfeuille F, Unoson C, Vogel J, & Wagner EG (2007) An antisense RNA inhibits translation by competing with standby ribosomes. *Mol Cell* 26(3):381-392.
45. Pfeiffer V, Papenfort K, Lucchini S, Hinton JC, & Vogel J (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol* 16(8):840-846.
46. Frohlich KS, Papenfort K, Berger AA, & Vogel J (2012) A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. *Nucleic Acids Res* 40(8):3623-3640.
47. Papenfort K & Vanderpool CK (2015) Target activation by regulatory RNAs in bacteria. *FEMS Microbiol Rev* 39(3):362-378.
48. Prevost K, *et al.* (2007) The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Mol Microbiol* 64(5):1260-1273.
49. Urban JH & Vogel J (2008) Two seemingly homologous noncoding RNAs act hierarchically to activate *glmS* mRNA translation. *PLoS Biol* 6(3):e64.
50. Quereda JJ, Ortega AD, Pucciarelli MG, & Garcia-Del Portillo F (2014) The *Listeria* Small RNA Rli27 Regulates a Cell Wall Protein inside Eukaryotic Cells by Targeting a Long 5'-UTR Variant. *PLoS Genet* 10(10):e1004765.
51. Morfeldt E, Taylor D, von Gabain A, & Arvidson S (1995) Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J* 14(18):4569-4577.
52. Frohlich KS, Papenfort K, Fekete A, & Vogel J (2013) A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J* 32(22):2963-2979.
53. Papenfort K, Sun Y, Miyakoshi M, Vanderpool CK, & Vogel J (2013) Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* 153(2):426-437.
54. Hammer BK & Bassler BL (2007) Regulatory small RNAs circumvent the conventional quorum sensing pathway in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 104(27):11145-11149.
55. Sharma CM, *et al.* (2011) Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* 81(5):1144-1165.
56. Masse E & Gottesman S (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* 99(7):4620-4625.
57. Masse E, Vanderpool CK, & Gottesman S (2005) Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol* 187(20):6962-6971.
58. Papenfort K, *et al.* (2006) SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. *Mol Microbiol* 62(6):1674-1688.
59. Johansen J, Rasmussen AA, Overgaard M, & Valentin-Hansen P (2006) Conserved small non-coding RNAs that belong to the sigma E regulon: role in down-regulation of outer membrane proteins. *J Mol Biol* 364(1):1-8.
60. Beisel CL & Storz G (2011) The base-pairing RNA Spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. *Mol Cell* 41(3):286-297.
61. Herzog R, Peschek N, Frohlich KS, Schumacher K, & Papenfort K (2019) Three autoinducer molecules act in concert to control virulence gene expression in *Vibrio cholerae*. *Nucleic Acids Res* 47(6):3171-3183.
62. Peschek N, Hoyos M, Herzog R, Forstner KU, & Papenfort K (2019) A conserved RNA seed-pairing domain directs small RNA-mediated stress resistance in enterobacteria. *EMBO J* 38(16):e101650.
63. 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. *Proc Natl Acad Sci U S A* 95(21):12462-12467.

## REFERENCES

64. Majdalani N, Hernandez D, & Gottesman S (2002) Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol Microbiol* 46(3):813-826.
65. Mandin P & Gottesman S (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO J* 29(18):3094-3107.
66. Battesti A, Majdalani N, & Gottesman S (2011) The RpoS-mediated general stress response in *Escherichia coli*. *Annu Rev Microbiol* 65:189-213.
67. Miyakoshi M, Chao Y, & Vogel J (2015) Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J* 34(11):1478-1492.
68. Melamed S, Adams PP, Zhang A, Zhang H, & Storz G (2020) RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles. *Mol Cell* 77(2):411-425 e417.
69. Gottesman S & Storz G (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* 3(12).
70. Holmqvist E & Wagner EGH (2017) Impact of bacterial sRNAs in stress responses. *Biochem Soc Trans* 45(6):1203-1212.
71. Westermann AJ (2018) Regulatory RNAs in Virulence and Host-Microbe Interactions. *Microbiol Spectr* 6(4).
72. Svenningsen SL (2018) Small RNA-Based Regulation of Bacterial Quorum Sensing and Biofilm Formation. *Microbiol Spectr* 6(4).
73. Beisel CL & Storz G (2010) Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol Rev* 34(5):866-882.
74. Nitzan M, Rehani R, & Margalit H (2017) Integration of Bacterial Small RNAs in Regulatory Networks. *Annu Rev Biophys* 46:131-148.
75. Brosse A & Guillier M (2018) Bacterial Small RNAs in Mixed Regulatory Networks. *Microbiol Spectr* 6(3).
76. Polayes DA, Rice PW, Garner MM, & Dahlberg JE (1988) Cyclic AMP-cyclic AMP receptor protein as a repressor of transcription of the *spf* gene of *Escherichia coli*. *J Bacteriol* 170(7):3110-3114.
77. Beisel CL & Storz G (2011) Discriminating tastes: physiological contributions of the Hfq-binding small RNA Spot 42 to catabolite repression. *RNA Biol* 8(5):766-770.
78. Durica-Mitic S, Gopel Y, & Gorke B (2018) Carbohydrate Utilization in Bacteria: Making the Most Out of Sugars with the Help of Small Regulatory RNAs. *Microbiol Spectr* 6(2).
79. Pappenfort K & Bassler BL (2016) Quorum sensing signal-response systems in Gram-negative bacteria. *Nat Rev Microbiol* 14(9):576-588.
80. Lenz DH, *et al.* (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* 118(1):69-82.
81. Miller MB, Skorupski K, Lenz DH, Taylor RK, & Bassler BL (2002) Parallel quorum sensing systems converge to regulate virulence in *Vibrio cholerae*. *Cell* 110(3):303-314.
82. Rutherford ST, van Kessel JC, Shao Y, & Bassler BL (2011) AphA and LuxR/HapR reciprocally control quorum sensing in *vibrios*. *Genes Dev* 25(4):397-408.
83. Zhu J, *et al.* (2002) Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 99(5):3129-3134.
84. Svenningsen SL, Waters CM, & Bassler BL (2008) A negative feedback loop involving small RNAs accelerates *Vibrio cholerae*'s transition out of quorum-sensing mode. *Genes Dev* 22(2):226-238.
85. Svenningsen SL, Tu KC, & Bassler BL (2009) Gene dosage compensation calibrates four regulatory RNAs to control *Vibrio cholerae* quorum sensing. *EMBO J* 28(4):429-439.
86. Hor J, Matera G, Vogel J, Gottesman S, & Storz G (2020) Trans-Acting Small RNAs and Their Effects on Gene Expression in *Escherichia coli* and *Salmonella enterica*. *EcoSal Plus* 9(1).

87. Feng L, *et al.* (2015) A *qrr* noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell* 160(1-2):228-240.
88. Overgaard M, Johansen J, Moller-Jensen J, & Valentin-Hansen P (2009) Switching off small RNA regulation with trap-mRNA. *Mol Microbiol* 73(5):790-800.
89. Franze de Fernandez MT, Eoyang L, & August JT (1968) Factor fraction required for the synthesis of bacteriophage Q $\beta$ -RNA. *Nature* 219(5154):588-590.
90. Valentin-Hansen P, Eriksen M, & Udesen C (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol Microbiol* 51(6):1525-1533.
91. Sun X, Zhulin I, & Wartell RM (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res* 30(17):3662-3671.
92. Vogel J & Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* 9(8):578-589.
93. Mura C, Randolph PS, Patterson J, & Cozen AE (2013) Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. *RNA Biol* 10(4):636-651.
94. Moller T, *et al.* (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell* 9(1):23-30.
95. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, & Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J* 21(13):3546-3556.
96. Sauer E & Weichenrieder O (2011) Structural basis for RNA 3'-end recognition by Hfq. *Proc Natl Acad Sci U S A* 108(32):13065-13070.
97. Otaka H, Ishikawa H, Morita T, & Aiba H (2011) PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc Natl Acad Sci U S A* 108(32):13059-13064.
98. Schu DJ, Zhang A, Gottesman S, & Storz G (2015) Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J* 34(20):2557-2573.
99. Link TM, Valentin-Hansen P, & Brennan RG (2009) Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci U S A* 106(46):19292-19297.
100. Robinson KE, Orans J, Kovach AR, Link TM, & Brennan RG (2014) Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Res* 42(4):2736-2749.
101. Horstmann N, Orans J, Valentin-Hansen P, Shelburne SA, 3rd, & Brennan RG (2012) Structural mechanism of *Staphylococcus aureus* Hfq binding to an RNA A-tract. *Nucleic Acids Res* 40(21):11023-11035.
102. Someya T, *et al.* (2012) Crystal structure of Hfq from *Bacillus subtilis* in complex with SELEX-derived RNA aptamer: insight into RNA-binding properties of bacterial Hfq. *Nucleic Acids Res* 40(4):1856-1867.
103. Wang L, *et al.* (2015) Structural insights into the recognition of the internal A-rich linker from OxyS sRNA by *Escherichia coli* Hfq. *Nucleic Acids Res* 43(4):2400-2411.
104. Panja S & Woodson SA (2012) Hfq proximity and orientation controls RNA annealing. *Nucleic Acids Res* 40(17):8690-8697.
105. Peng Y, Soper TJ, & Woodson SA (2014) Positional effects of AAN motifs in *rpoS* regulation by sRNAs and Hfq. *J Mol Biol* 426(2):275-285.
106. Sauer E, Schmidt S, & Weichenrieder O (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc Natl Acad Sci U S A* 109(24):9396-9401.
107. Zhang A, Schu DJ, Tjaden BC, Storz G, & Gottesman S (2013) Mutations in interaction surfaces differentially impact *E. coli* Hfq association with small RNAs and their mRNA targets. *J Mol Biol* 425(19):3678-3697.
108. Dimastrogiovanni D, *et al.* (2014) Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *Elife* 3.

## REFERENCES

109. Panja S, Schu DJ, & Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. *Nucleic Acids Res* 41(15):7536-7546.
110. Vincent HA, *et al.* (2012) Characterization of *Vibrio cholerae* Hfq provides novel insights into the role of the Hfq C-terminal region. *J Mol Biol* 420(1-2):56-69.
111. Santiago-Frangos A, Kavita K, Schu DJ, Gottesman S, & Woodson SA (2016) C-terminal domain of the RNA chaperone Hfq drives sRNA competition and release of target RNA. *Proc Natl Acad Sci U S A* 113(41):E6089-E6096.
112. Morita T, Maki K, & Aiba H (2005) RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev* 19(18):2176-2186.
113. Prevost K, Desnoyers G, Jacques JF, Lavoie F, & Masse E (2011) Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes Dev* 25(4):385-396.
114. Gottesman S (2004) The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu Rev Microbiol* 58:303-328.
115. Masse E, Escorcia FE, & Gottesman S (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev* 17(19):2374-2383.
116. Hajnsdorf E & Regnier P (2000) Host factor Hfq of *Escherichia coli* stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proc Natl Acad Sci U S A* 97(4):1501-1505.
117. Mohanty BK, Maples VF, & Kushner SR (2004) The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*. *Mol Microbiol* 54(4):905-920.
118. Vytvytska O, Moll I, Kaberdin VR, von Gabain A, & Blasi U (2000) Hfq (HF1) stimulates *ompA* mRNA decay by interfering with ribosome binding. *Genes Dev* 14(9):1109-1118.
119. Updegrove TB, Zhang A, & Storz G (2016) Hfq: the flexible RNA matchmaker. *Curr Opin Microbiol* 30:133-138.
120. Holmqvist E & Vogel J (2018) RNA-binding proteins in bacteria. *Nat Rev Microbiol* 16(10):601-615.
121. Moon K & Gottesman S (2011) Competition among Hfq-binding small RNAs in *Escherichia coli*. *Mol Microbiol* 82(6):1545-1562.
122. Fender A, Elf J, Hampel K, Zimmermann B, & Wagner EG (2010) RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev* 24(23):2621-2626.
123. Wagner EG (2013) Cycling of RNAs on Hfq. *RNA Biol* 10(4):619-626.
124. Hussein R & Lim HN (2011) Disruption of small RNA signaling caused by competition for Hfq. *Proc Natl Acad Sci U S A* 108(3):1110-1115.
125. Faigenbaum-Romm R, *et al.* (2020) Hierarchy in Hfq Chaperon Occupancy of Small RNA Targets Plays a Major Role in Their Regulation. *Cell Rep* 30(9):3127-3138 e3126.
126. Melamed S, *et al.* (2016) Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol Cell* 63(5):884-897.
127. Sobrero P & Valverde C (2012) The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit Rev Microbiol* 38(4):276-299.
128. Tsui HC, Leung HC, & Winkler ME (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. *Mol Microbiol* 13(1):35-49.
129. Sonnleitner E, *et al.* (2003) Reduced virulence of a *hfq* mutant of *Pseudomonas aeruginosa* O1. *Microb Pathog* 35(5):217-228.
130. Ding Y, Davis BM, & Waldor MK (2004) Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol Microbiol* 53(1):345-354.
131. Sittka A, Pfeiffer V, Tedin K, & Vogel J (2007) The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol Microbiol* 63(1):193-217.

132. Romeo T, Gong M, Liu MY, & Brun-Zinkernagel AM (1993) Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J Bacteriol* 175(15):4744-4755.
133. Babitzke P & Romeo T (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol* 10(2):156-163.
134. Kunte HJ, Crane RA, Culham DE, Richmond D, & Wood JM (1999) Protein ProQ influences osmotic activation of compatible solute transporter ProP in *Escherichia coli* K-12. *J Bacteriol* 181(5):1537-1543.
135. Smirnov A, *et al.* (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci U S A* 113(41):11591-11596.
136. Holmqvist E, Li L, Bischler T, Barquist L, & Vogel J (2018) Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3' Ends. *Mol Cell* 70(5):971-982 e976.
137. Smirnov A, Wang C, Drewry LL, & Vogel J (2017) Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. *EMBO J* 36(8):1029-1045.
138. Jiang W, Hou Y, & Inouye M (1997) CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J Biol Chem* 272(1):196-202.
139. Gopel Y, Papenfort K, Reichenbach B, Vogel J, & Gorke B (2013) Targeted decay of a regulatory small RNA by an adaptor protein for RNase E and counteraction by an anti-adaptor RNA. *Genes Dev* 27(5):552-564.
140. Ait-Bara S, Carpousis AJ, & Quentin Y (2015) RNase E in the gamma-Proteobacteria: conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains. *Mol Genet Genomics* 290(3):847-862.
141. Callaghan AJ, *et al.* (2005) Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature* 437(7062):1187-1191.
142. Callaghan AJ, *et al.* (2004) Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E. *J Mol Biol* 340(5):965-979.
143. Carpousis AJ (2007) The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol* 61:71-87.
144. Marcaida MJ, DePristo MA, Chandran V, Carpousis AJ, & Luisi BF (2006) The RNA degradosome: life in the fast lane of adaptive molecular evolution. *Trends Biochem Sci* 31(7):359-365.
145. Mohanty BK & Kushner SR (2018) Enzymes Involved in Posttranscriptional RNA Metabolism in Gram-Negative Bacteria. *Microbiol Spectr* 6(2).
146. Li Z, Pandit S, & Deutscher MP (1999) RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. *EMBO J* 18(10):2878-2885.
147. Li Z & Deutscher MP (2002) RNase E plays an essential role in the maturation of *Escherichia coli* tRNA precursors. *RNA* 8(1):97-109.
148. Ow MC & Kushner SR (2002) Initiation of tRNA maturation by RNase E is essential for cell viability in *E. coli*. *Genes Dev* 16(9):1102-1115.
149. Dar D & Sorek R (2018) Bacterial Noncoding RNAs Excised from within Protein-Coding Transcripts. *mBio* 9(5).
150. McDowall KJ, Lin-Chao S, & Cohen SN (1994) A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J Biol Chem* 269(14):10790-10796.
151. McDowall KJ, Kaberdin VR, Wu SW, Cohen SN, & Lin-Chao S (1995) Site-specific RNase E cleavage of oligonucleotides and inhibition by stem-loops. *Nature* 374(6519):287-290.
152. Bouvier M & Carpousis AJ (2011) A tale of two mRNA degradation pathways mediated by RNase E. *Mol Microbiol* 82(6):1305-1310.

## REFERENCES

153. Mackie GA (1998) Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* 395(6703): 720-723.
154. Baker KE & Mackie GA (2003) Ectopic RNase E sites promote bypass of 5'-end-dependent mRNA decay in *Escherichia coli*. *Mol Microbiol* 47(1):75-88.
155. Clarke JE, Kime L, Romero AD, & McDowall KJ (2014) Direct entry by RNase E is a major pathway for the degradation and processing of RNA in *Escherichia coli*. *Nucleic Acids Res* 42(18):11733-11751.
156. Bandyra KJ, Wandzik JM, & Luisi BF (2018) Substrate Recognition and Autoinhibition in the Central Ribonuclease RNase E. *Mol Cell* 72(2):275-285 e274.
157. Papenfort K, Bouvier M, Mika F, Sharma CM, & Vogel J (2010) Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc Natl Acad Sci U S A* 107(47):20435-20440.
158. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136(2):215-233.
159. O'Brien J, Hayder H, Zayed Y, & Peng C (2018) Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol (Lausanne)* 9:402.
160. Irnov I, Sharma CM, Vogel J, & Winkler WC (2010) Identification of regulatory RNAs in *Bacillus subtilis*. *Nucleic Acids Res* 38(19):6637-6651.
161. Wilms I, Overloper A, Nowrousian M, Sharma CM, & Narberhaus F (2012) Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen *Agrobacterium tumefaciens*. *RNA Biol* 9(4):446-457.
162. Li SC, *et al.* (2010) Identification of homologous microRNAs in 56 animal genomes. *Genomics* 96(1):1-9.
163. Friedlander MR, *et al.* (2014) Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol* 15(4):R57.
164. Holmqvist E, *et al.* (2016) Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J* 35(9):991-1011.
165. Chi SW, Zang JB, Mele A, & Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460(7254):479-486.
166. Ule J, Jensen K, Mele A, & Darnell RB (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* 37(4):376-386.
167. Zhang C & Darnell RB (2011) Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* 29(7):607-614.
168. Melamed S, *et al.* (2018) Mapping the small RNA interactome in bacteria using RIL-seq. *Nat Protoc* 13(1):1-33.
169. 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.
170. Waters SA, *et al.* (2017) Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *EMBO J* 36(3):374-387.
171. Iosub IA, *et al.* (2020) Hfq CLASH uncovers sRNA-target interaction networks linked to nutrient availability adaptation. *Elife* 9.
172. Barquist L & Vogel J (2015) Accelerating Discovery and Functional Analysis of Small RNAs with New Technologies. *Annu Rev Genet* 49:367-394.
173. Reidl J & Klose KE (2002) *Vibrio cholerae* and cholera: out of the water and into the host. *FEMS Microbiol Rev* 26(2):125-139.
174. Charles RC & Ryan ET (2011) Cholera in the 21st century. *Curr Opin Infect Dis* 24(5):472-477.
175. Heidelberg JF, *et al.* (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406(6795):477-483.

176. Kojima S, Yamamoto K, Kawagishi I, & Homma M (1999) The polar flagellar motor of *Vibrio cholerae* is driven by an Na<sup>+</sup> motive force. *J Bacteriol* 181(6):1927-1930.
177. Teschler JK, *et al.* (2015) Living in the matrix: assembly and control of *Vibrio cholerae* biofilms. *Nat Rev Microbiol* 13(5):255-268.
178. Donlan RM (2002) Biofilms: microbial life on surfaces. *Emerg Infect Dis* 8(9):881-890.
179. Donlan RM & Costerton JW (2002) Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin Microbiol Rev* 15(2):167-193.
180. Meibom KL, *et al.* (2004) The *Vibrio cholerae* chitin utilization program. *Proc Natl Acad Sci U S A* 101(8):2524-2529.
181. Meibom KL, Blokesch M, Dolganov NA, Wu CY, & Schoolnik GK (2005) Chitin induces natural competence in *Vibrio cholerae*. *Science* 310(5755):1824-1827.
182. Silva AJ & Benitez JA (2016) *Vibrio cholerae* Biofilms and Cholera Pathogenesis. *PLoS Negl Trop Dis* 10(2):e0004330.
183. Zhu J & Mekalanos JJ (2003) Quorum sensing-dependent biofilms enhance colonization in *Vibrio cholerae*. *Dev Cell* 5(4):647-656.
184. Taylor RK, Miller VL, Furlong DB, & Mekalanos JJ (1987) Use of *phoA* gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc Natl Acad Sci U S A* 84(9):2833-2837.
185. Herrington DA, *et al.* (1988) Toxin, toxin-coregulated pili, and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. *J Exp Med* 168(4):1487-1492.
186. Karaolis DK, *et al.* (1998) A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proc Natl Acad Sci U S A* 95(6):3134-3139.
187. Kaper JB, Morris JG, Jr., & Levine MM (1995) Cholera. *Clin Microbiol Rev* 8(1):48-86.
188. Morris JG, Jr. (2011) Cholera--modern pandemic disease of ancient lineage. *Emerg Infect Dis* 17(11):2099-2104.
189. Bardill JP & Hammer BK (2012) Non-coding sRNAs regulate virulence in the bacterial pathogen *Vibrio cholerae*. *RNA Biol* 9(4):392-401.
190. Perez-Reytor D, *et al.* (2016) Role of Non-coding Regulatory RNA in the Virulence of Human Pathogenic *Vibrios*. *Front Microbiol* 7:2160.
191. Song T, *et al.* (2008) A new *Vibrio cholerae* sRNA modulates colonization and affects release of outer membrane vesicles. *Mol Microbiol* 70(1):100-111.
192. Richard AL, Withey JH, Beyhan S, Yildiz F, & DiRita VJ (2010) The *Vibrio cholerae* virulence regulatory cascade controls glucose uptake through activation of TarA, a small regulatory RNA. *Mol Microbiol* 78(5):1171-1181.
193. Bradley ES, Bodi K, Ismail AM, & Camilli A (2011) A genome-wide approach to discovery of small RNAs involved in regulation of virulence in *Vibrio cholerae*. *PLoS Pathog* 7(7):e1002126.
194. Pulvermacher SC, Stauffer LT, & Stauffer GV (2008) The role of the small regulatory RNA GcvB in GcvB/mRNA posttranscriptional regulation of *oppA* and *dppA* in *Escherichia coli*. *FEMS Microbiol Lett* 281(1):42-50.
195. Davis BM, Quinones M, Pratt J, Ding Y, & Waldor MK (2005) Characterization of the small untranslated RNA RyhB and its regulon in *Vibrio cholerae*. *J Bacteriol* 187(12):4005-4014.
196. Mey AR, Craig SA, & Payne SM (2005) Characterization of *Vibrio cholerae* RyhB: the RyhB regulon and role of *ryhB* in biofilm formation. *Infect Immun* 73(9):5706-5719.
197. Liu JM, *et al.* (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res* 37(6):e46.
198. Zhang A, *et al.* (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol* 50(4):1111-1124.

## REFERENCES

199. Forst S, Dowds B, Boemare N, & Stackebrandt E (1997) *Xenorhabdus* and *Photorhabdus* spp.: bugs that kill bugs. *Annu Rev Microbiol* 51:47-72.
200. Clarke DJ (2008) *Photorhabdus*: a model for the analysis of pathogenicity and mutualism. *Cell Microbiol* 10(11):2159-2167.
201. Bode HB (2009) Entomopathogenic bacteria as a source of secondary metabolites. *Curr Opin Chem Biol* 13(2):224-230.
202. Challinor VL & Bode HB (2015) Bioactive natural products from novel microbial sources. *Ann N Y Acad Sci* 1354:82-97.
203. Tobias NJ, *et al.* (2017) *Photorhabdus*-nematode symbiosis is dependent on *hfq*-mediated regulation of secondary metabolites. *Environ Microbiol* 19(1):119-129.
204. Saliba AE, S CS, & Vogel J (2017) New RNA-seq approaches for the study of bacterial pathogens. *Curr Opin Microbiol* 35:78-87.
205. Bilusic I, Popitsch N, Rescheneder P, Schroeder R, & Lybecker M (2014) Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. *RNA Biol* 11(5):641-654.
206. Davis BM & Waldor MK (2007) RNase E-dependent processing stabilizes MicX, a *Vibrio cholerae* sRNA. *Mol Microbiol* 65(2):373-385.
207. Yamamoto S, *et al.* (2011) Identification of a chitin-induced small RNA that regulates translation of the *tfoX* gene, encoding a positive regulator of natural competence in *Vibrio cholerae*. *J Bacteriol* 193(8):1953-1965.
208. Dugar G, *et al.* (2016) The CsrA-FliW network controls polar localization of the dual-function flagellin mRNA in *Campylobacter jejuni*. *Nat Commun* 7:11667.
209. Lalaouna D, *et al.* (2015) A 3' external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Mol Cell* 58(3):393-405.
210. Lalaouna D, Prevost K, Eyraud A, & Masse E (2017) Identification of unknown RNA partners using MAPS. *Methods* 117:28-34.
211. Corcoran CP, *et al.* (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* 84(3):428-445.
212. Wright PR, *et al.* (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res* 42:W119-123.
213. Georg J, *et al.* (2020) The power of cooperation: Experimental and computational approaches in the functional characterization of bacterial sRNAs. *Mol Microbiol* 113(3):603-612.
214. Beyer D, Skripkin E, Wadzack J, & Nierhaus KH (1994) How the ribosome moves along the mRNA during protein synthesis. *J Biol Chem* 269(48):30713-30717.
215. Huttenhofer A & Noller HF (1994) Footprinting mRNA-ribosome complexes with chemical probes. *EMBO J* 13(16):3892-3901.
216. Bouvier M, Sharma CM, Mika F, Nierhaus KH, & Vogel J (2008) Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol Cell* 32(6):827-837.
217. Desnoyers G & Masse E (2012) Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes Dev* 26(7):726-739.
218. de Smit MH & van Duin J (2003) Translational standby sites: how ribosomes may deal with the rapid folding kinetics of mRNA. *J Mol Biol* 331(4):737-743.
219. Unoson C & Wagner EG (2007) Dealing with stable structures at ribosome binding sites: bacterial translation and ribosome standby. *RNA Biol* 4(3):113-117.
220. Aiba H (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr Opin Microbiol* 10(2):134-139.
221. Moller T, Franch T, Udesen C, Gerdes K, & Valentin-Hansen P (2002) Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev* 16(13):1696-1706.

222. Updegrove TB, Shabalina SA, & Storz G (2015) How do base-pairing small RNAs evolve? *FEMS Microbiol Rev* 39(3):379-391.
223. Pappenfort K, Podkaminski D, Hinton JC, & Vogel J (2012) The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. *Proc Natl Acad Sci U S A* 109(13):E757-764.
224. Loh E, *et al.* (2009) A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell* 139(4):770-779.
225. Durand S & Condon C (2018) RNases and Helicases in Gram-Positive Bacteria. *Microbiol Spectr* 6(2).
226. Bouloc P & Repoila F (2016) Fresh layers of RNA-mediated regulation in Gram-positive bacteria. *Curr Opin Microbiol* 30:30-35.
227. Nielsen JS, *et al.* (2010) Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in *Listeria monocytogenes*. *Nucleic Acids Res* 38(3):907-919.
228. Nicolas P, *et al.* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 335(6072):1103-1106.
229. Mars RA, Nicolas P, Denham EL, & van Dijl JM (2016) Regulatory RNAs in *Bacillus subtilis*: a Gram-Positive Perspective on Bacterial RNA-Mediated Regulation of Gene Expression. *Microbiol Mol Biol Rev* 80(4):1029-1057.
230. Lasa I, *et al.* (2011) Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc Natl Acad Sci U S A* 108(50):20172-20177.
231. Toledo-Arana A, *et al.* (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* 459(7249):950-956.
232. Rasmussen S, Nielsen HB, & Jarmer H (2009) The transcriptionally active regions in the genome of *Bacillus subtilis*. *Mol Microbiol* 73(6):1043-1057.
233. Peer A & Margalit H (2014) Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. *RNA* 20(7):994-1003.
234. Baekkedal C & Haugen P (2015) The Spot 42 RNA: A regulatory small RNA with roles in the central metabolism. *RNA Biol* 12(10):1071-1077.
235. Richter AS & Backofen R (2012) Accessibility and conservation: general features of bacterial small RNA-mRNA interactions? *RNA Biol* 9(7):954-965.
236. Keswani C, *et al.* (2020) Antimicrobial secondary metabolites from agriculturally important bacteria as next-generation pesticides. *Appl Microbiol Biotechnol* 104(3):1013-1034.
237. Barnard AM, *et al.* (2007) Quorum sensing, virulence and secondary metabolite production in plant soft-rotting bacteria. *Philos Trans R Soc Lond B Biol Sci* 362(1483):1165-1183.
238. Harris SJ, Shih YL, Bentley SD, & Salmond GP (1998) The *hexA* gene of *Erwinia carotovora* encodes a LysR homologue and regulates motility and the expression of multiple virulence determinants. *Mol Microbiol* 28(4):705-717.
239. Mukherjee A, Cui Y, Ma W, Liu Y, & Chatterjee AK (2000) *hexA* of *Erwinia carotovora* ssp. *carotovora* strain Ecc71 negatively regulates production of RpoS and *rsmB* RNA, a global regulator of extracellular proteins, plant virulence and the quorum-sensing signal, N-(3-oxohexanoyl)-L-homoserine lactone. *Environ Microbiol* 2(2):203-215.
240. Zeng Q, McNally RR, & Sundin GW (2013) Global small RNA chaperone Hfq and regulatory small RNAs are important virulence regulators in *Erwinia amylovora*. *J Bacteriol* 195(8):1706-1717.
241. Matilla MA, Leeper FJ, & Salmond GP (2015) Biosynthesis of the antifungal haterumalide, oocydin A, in *Serratia*, and its regulation by quorum sensing, RpoS and Hfq. *Environ Microbiol* 17(8):2993-3008.
242. Wilf NM & Salmond GPC (2012) The stationary phase sigma factor, RpoS, regulates the production of a carbapenem antibiotic, a bioactive prodigiosin and virulence in the enterobacterial pathogen *Serratia* sp. ATCC 39006. *Microbiology* 158(Pt 3):648-658.

## REFERENCES

243. Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* 8(6):450-461.
244. Rosenfeld N, Elowitz MB, & Alon U (2002) Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* 323(5):785-793.
245. 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 Res* 36(21):6781-6794.
246. Shen-Orr SS, Milo R, Mangan S, & Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 31(1):64-68.
247. Mangan S & Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A* 100(21):11980-11985.
248. Mangan S, Zaslaver A, & Alon U (2003) The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol* 334(2):197-204.
249. Thomason MK, Fontaine F, De Lay N, & Storz G (2012) A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol Microbiol* 84(1):17-35.
250. Boisset S, *et al.* (2007) *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev* 21(11):1353-1366.
251. Nitzan M, *et al.* (2015) A defense-offense multi-layered regulatory switch in a pathogenic bacterium. *Nucleic Acids Res* 43(3):1357-1369.
252. Coyer J, Andersen J, Forst SA, Inouye M, & Delihans N (1990) MicF RNA in *ompB* mutants of *Escherichia coli*: different pathways regulate MicF RNA levels in response to osmolarity and temperature change. *J Bacteriol* 172(8):4143-4150.
253. Shimoni Y, *et al.* (2007) Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Syst Biol* 3:138.
254. Chen S, Zhang A, Blyn LB, & Storz G (2004) MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J Bacteriol* 186(20):6689-6697.
255. Zhang YM & Rock CO (2009) Transcriptional regulation in bacterial membrane lipid synthesis. *J Lipid Res* 50 Suppl:S115-119.
256. Shi W, *et al.* (2015) The 40-residue insertion in *Vibrio cholerae* FadR facilitates binding of an additional fatty acyl-CoA ligand. *Nat Commun* 6:6032.
257. Campbell JW & Cronan JE, Jr. (2001) *Escherichia coli* FadR positively regulates transcription of the *fabB* fatty acid biosynthetic gene. *J Bacteriol* 183(20):5982-5990.
258. Campbell JW & Cronan JE, Jr. (2002) The enigmatic *Escherichia coli* *fadE* gene is *yafH*. *J Bacteriol* 184(13):3759-3764.
259. DiRusso CC, Heimert TL, & Metzger AK (1992) Characterization of FadR, a global transcriptional regulator of fatty acid metabolism in *Escherichia coli*. Interaction with the *fadB* promoter is prevented by long chain fatty acyl coenzyme A. *J Biol Chem* 267(12):8685-8691.
260. Fujita Y, Matsuoka H, & Hirooka K (2007) Regulation of fatty acid metabolism in bacteria. *Mol Microbiol* 66(4):829-839.
261. Iram SH & Cronan JE (2005) Unexpected functional diversity among FadR fatty acid transcriptional regulatory proteins. *J Biol Chem* 280(37):32148-32156.
262. Rivera-Chavez F & Mekalanos JJ (2019) Cholera toxin promotes pathogen acquisition of host-derived nutrients. *Nature* 572(7768):244-248.
263. Giles DK, Hankins JV, Guan Z, & Trent MS (2011) Remodelling of the *Vibrio cholerae* membrane by incorporation of exogenous fatty acids from host and aquatic environments. *Mol Microbiol* 79(3):716-728.
264. Mandlik A, *et al.* (2011) RNA-Seq-based monitoring of infection-linked changes in *Vibrio cholerae* gene expression. *Cell Host Microbe* 10(2):165-174.

265. Yang S, *et al.* (2020) *Vibrio cholerae* VC1741 (PsrA) enhances the colonization of the pathogen in infant mice intestines in the presence of the long-chain fatty acid, oleic acid. *Microb Pathog* 147:104443.

# Abbreviations

AGO	Argonaute
CDS	coding sequence
CLASH	cross-linking, ligation and sequencing of hybrids
CLIP-seq	crosslinking and immunoprecipitation followed by RNA-sequencing
CopraRNA	Comparative prediction algorithm for small RNA targets
CRP	cAMP receptor protein
DNA	deoxyribonucleic acid
EMSA	electrophoretic mobility shift assay
FA	fatty acid
FFL	feed forward loop
gDNA	genomic DNA
GEO	gene expression omnibus
GFP	green fluorescent protein
Grad-seq	gradient profiling by sequencing
HCD	high cell density
HITS-CLIP	high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation
IGR	intergenic region
IP	immunoprecipitation
LB	Lennox broth
LCD	low cell density
LCFA	long chain fatty acid
MAPS	MS2-affinity purification coupled with RNA-sequencing
miRNA	microRNA
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
OD	optical density
PAA	polyacrylamide
P:C:I	phenol:chloroform:isoamylalcohol
PCR	polymerase chain reaction
PBS	phosphate-buffered saline
QS	quorum sensing

RBP	RNA-binding protein
RBS	ribosome binding site
RIL-seq	RNA interaction by ligation and sequencing
RIP-seq	RNA co-immunoprecipitation followed by RNA-sequencing
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNase	ribonuclease
RNA-seq	RNA high-throughput sequencing
RNAP	RNA polymerase
RNAT	RNA thermometer
rRNA	ribosomal RNA
SD	standard deviation
SD sequence	Shine-Dalgarno sequence
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	standard error of the mean
SOC	Super Optimal Broth (SOB) with catabolite repression
sRNA	small RNA
SSC	saline-sodium citrate
TIER-seq	transiently inactivating an endoribonuclease followed by RNA-seq
TIR	translation initiation region
tRNA	transfer RNA
UTR	untranslated region

# Acknowledgements

Out of the many people that accompanied me on this journey, I would first and foremost like to thank my supervisor Prof. Dr. Kai Papenfort for giving me the opportunity to do my PhD in his group and for introducing me to the fascinating world of regulatory RNAs. I am deeply grateful for the constant scientific advice and support throughout the years. I would also like to thank him for giving me the opportunity to attend great scientific conferences and to go to the lab of Dr. Gisela Storz at the National Institutes of Health for an amazing research stay.

I would like to express my sincere gratitude to Prof. Dr. Simon Ringgaard to be the second examiner of this thesis, as well as to PD Dr. Tatjana Kleine, Prof. Dr. Christof Osman, Prof. Dr. Nicolas Gompel and Prof. Dr. Peter Geigenberger to be part of the committee. Furthermore, I am very thankful to my TAC members PD Dr. Jörg Meurer and Prof. Dr. Jörg Nickelsen for supporting my projects and giving valuable input throughout the years.

This thesis would not be the same without great collaboration partners. Thank you to Dr. Sahar Melamed and Dr. Gisela Storz, and all members of the Storz lab, for great scientific input and more than a nice welcome to NIH. Thank you as well to Dr. Nick Neubacher and Prof. Dr. Helge Bode from the Goethe University Frankfurt for an exciting cooperation on *Photorhabdus* sRNAs.

Further, I would like to thank Dr. Kathrin Fröhlich for valuable scientific ideas, and all former and current members of the Papenfort group for insightful discussions and an inspiring atmosphere. Special thanks to my colleagues Mona, Kavyaa and Yannik for the constant support, warmth and positive atmosphere. I couldn't imagine better colleagues than you! A big thank you also to my student Jessy for her reliable and enormous help with many experiments.

I also would like to thank the coordinators of the LSM Graduate School Francisca Rosa Mende and Nadine Hamze for their excellent organization of all LSM events.

Last but not least, thank you to my family for the endless support and to my wonderful husband Raphael for his constant encouragement and help whenever needed.

# Curriculum vitae

Removed for protection of privacy.

Aus Datenschutzgründen entfernt.