Identification of Small Regulatory RNAs and Their Targets in Bacteria

Dissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) der Technischen Fakultät der Universität Bielefeld

vorgelegt von

Cynthia Mira Sharma

Bielefeld, im Mai 2009

Dissertation zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften (Dr. rer. nat.)
der Technischen Fakultät der Universität Bielefeld.

Vorgelegt von:	Dipl.-Biol. Cynthia Mira Sharma
Angefertigt im:	Max-Planck-Institut für Infektionsbiologie, Berlin
Tag der Einreichung:	04. Mai 2009
Tag der Verteidigung:	22. Juli 2009

Gutachter:

Prof. Dr. Robert Giegerich, Universität Bielefeld
Dr. Jörg Vogel, Max-Planck-Institut für Infektionsbiologie, Berlin
Prof. Dr. Wolfgang Hess, Universität Freiburg

Prüfungsausschuss:

Prof. Dr. Jens Stoye, Universität Bielefeld
Prof. Dr. Robert Giegerich, Universität Bielefeld
Dr. Jörg Vogel, Max-Planck-Institut für Infektionsbiologie, Berlin
Prof. Dr. Wolfgang Hess, Universität Freiburg
Dr. Marilia Braga, Universität Bielefeld

Gedruckt auf alterungsbeständigem Papier nach ISO 9706.

Dedicated to my parents, Ingrid Sharma and Som Deo Sharma

ACKNOWLEDGEMENTS

At this point, I would like to thank everybody who contributed in whatever form to this thesis. In particular, I am grateful to the following persons:

… my supervisor Dr. Jörg Vogel for giving me the opportunity to work in a superb scientific environment and for his continuous support and guidance throughout the past several years,

… Prof. Dr. Robert Giegerich for enabling me to submit this thesis to the University of Bielefeld and for valuable suggestions during several visits to his institute,

… Prof. Dr. Wolfgang Hess for his willingness to evaluate this thesis,

… Dr. Fabien Darfeuille (a. k. a. the “master of the gels”) for a great stay in his lab in Bordeaux and many fruitful discussions,

… Dr. Steve Hoffmann, Kai Papenfort, Dr. Titia Plantinga, and Dr. Alexandra Sittka for their contributions to this thesis, helpful discussions, and enjoyable collaborations,

… Dr. Marc Rehmsmeier for initial suggestions about sRNA target predictions using `RNAhybrid`,

… Franziska “Franntz” Seifert and Sandy Pernitzsch for technical assistance in various experiments,

… Birgit Grett, unserer „guten Fee“ im Labor, für ihren Beitrag zum Gelingen unserer Experimente,

… all proofreaders of this thesis, especially Kathrin “Kathi” Fröhlich for her final scrutiny and Kai “Lillifee” Papenfort for his critical comments and, in addition, for a lot of fun with them in the lab, their support and help,

… all my colleagues and former members of the Vogel lab for a fantastic time and kind cooperation,

… last but not least my parents, my sister, and my friends for their patience and encouragement,

… and, of course, Nikolaus for immensely supporting and encouraging me, especially during the past few months.

CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

(v/v)	(volume/volume)
(w/v)	(weight/volume in g/ml)
A	adenine
aa	amino acid
Amp	ampicillin
APS	ammonium persulfate
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	base pair
Cat	chloramphenicol cassette
Cm	chloramphenicol
CDS	coding sequence
C	cytosine
CTP	cytidine triphosphate
cDNA	complementary DNA
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetate
Fig.	Figure
G	guanine
GFP	green fluorescent protein
GTP	guanosine triphosphate
Hfq	host factor for Qβ replication
IGR	intergenic region
Kan	kanamycin
kcal	kilocalorie
LB	Lennox broth
MB	megabyte
mfe structure	minimum-free-energy structure
mRNA	messenger RNA
miRNA	microRNA
NCBI	National Center for Biotechnology Information
ncRNA	noncoding RNA
nt	nucleotide(s)
OD	optical density
ORF	open reading frame
PAA	polyacrylamide

PBS	phosphate buffered saline
PCR	polymerase chain reaction
Perl	Practical Extraction and Report Language
RACE	rapid amplification of cDNA ends
RBS	ribosome binding site
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal RNA
rNTP	ribonucleotide
sRNA	small RNA
siRNA	short-interfering RNA
SD	Shine-Dalgarno
SDS	sodiumdodecylsulfate
SRP	signal recognition particle
T	thymine
TBE	Tris/Borate/EDTA
TEMED	tetramethylethylendiamin
TIR	translation initiation region
Tris	tris-(hydroxymethyl)-aminomethan
tRNA	transfer RNA
U	uracil
UTP	uridine triphosphate
UTR	untranslated region

UNITS

°C	degree Celsius
Da	Dalton
g	gram
h	hour
l	litre
M	molar
min	minute
molar	gram molecule
rpm	rounds per minute
s	second
u	unit
V	Volt
W	Watt

MULTIPLES

M	mega (10^6)
k	kilo (10^3)
m	milli (10^{-3})
μ	micro(10^{-6})
n	nano (10^{-9})
p	pico (10^{-12})

CHAPTER 1

INTRODUCTION

RNA is a versatile molecule and due to its wide range of biochemical properties it is capable of multifarious functions. The linear sequence of RNA makes it a simple source of genetic information, whereas the property to form secondary and tertiary structures allows its interaction with other macromolecules and provides environments for catalytic activities. Thus, besides the role of RNA molecules as information-carrying intermediaries in gene expression, they act as key catalytic, structural, and regulatory elements in the cell. In bacteria, the discovery of a staggering number of small regulatory RNAs (sRNAs) by systematic searches of sequenced genomes over the last years led to an increasing recognition of the potential impact of sRNAs on bacterial physiology (for a review see, *e. g.*, Waters & Storz, 2009). These sRNAs act as post-transcriptional regulators of bacterial gene expression in response to diverse growth and environmental stress conditions. In contrast to *cis*-encoded antisense RNAs of mobile elements such as plasmids, the majority of bacterial sRNAs seems to bind by imperfect basepairing to *trans*-encoded mRNAs and thereby inhibit translation or lead to mRNA degradation. The early studies have often focused on interactions with single target mRNAs, but there is growing evidence that sRNAs can regulate many diverse mRNAs in parallel (reviewed in Papenfort & Vogel, 2009, *in press*). However, the understanding of how sRNAs could directly control multiple mRNAs by antisense mechanism has been limited by the low number of validated sRNA-target interactions and, hence, the difficulty to reliably predict new interactions.

In this thesis, two aspects of sRNA-mediated regulation in bacteria are investigated: (1) multiple target regulation and (2) approaches for the identification of novel sRNAs in bacteria. The first part addresses the question how sRNA targets can be identified and how multiple targets can be directly regulated by one sRNA. For this purpose, biocomputational and experimental approaches for the identification and validation of targets of a small RNA, GcvB, from *Salmonella enterica* serovar Typhimurium (from here on *Salmonella*) are presented. Furthermore, it is shown that a conserved region within GcvB RNA directly interacts with multiple mRNAs of genes involved in amino acid transport and biosynthesis. It is shown how the identification of this conserved element can be used to refine experimental and biocomputational target-identification approaches.

The second part deals with the identification of novel sRNAs. Bioinformatics-based approaches often rely on the prediction of orphan transcription signals and primary sequence conservation of sRNA candidates within closely related species or on the conservation of RNA structure (reviewed in Livny & Waldor, 2007). This implies the availability of related genome sequences and well-defined promoter and terminator models. In contrast, approaches based on shotgun-cloning and direct sequencing of RNA (so-called RNomics) allow to identify novel sRNAs without prior knowledge but were, until recently, limited by the cost-intensive Sanger sequencing (Hüttenhofer & Vogel,

2006). In this thesis, the use of high-throughput sequencing for the identification of sRNAs bound to the RNA-binding protein Hfq in *Salmonella* is demonstrated. Furthermore, deep sequencing reveals the primary transcriptome of the major human pathogen *Helicobacter pylori*, a bacterium in which no sRNAs have been described. Moreover, an approach based on selective sequencing of cDNA libraries specifically enriched for primary transcripts is developed which allows to define a global map of transcriptional start sites of mRNAs in *H. pylori*.

Organization of this thesis

First, the biological background of regulation by small RNAs in bacteria is reviewed in Chapter 2 with the main emphasis on strategies for the identification of sRNAs and their targets in bacteria.

In Chapters 3 and 4, multiple target regulation by bacterial sRNAs is investigated by functional characterization of GcvB RNA from *Salmonella*. Specifically, Chapter 3 describes the proteomics and bioinformatics-based identification of seven ABC transporter mRNAs as GcvB targets. Analysis of target gene fusion regulation *in vivo*, as well as *in vitro* structure probing and translation assays show that *Salmonella* GcvB sRNA directly binds with a conserved G/U-rich region to extended C/A rich elements of these target mRNAs. The identified target sites are located inside or upstream of the ribosome binding site and may also serve as translational enhancer elements. This suggests mRNA regions distant from Shine-Dalgarno (SD) sequences and highly conserved regions in sRNAs as important elements for the identification of sRNA targets.

In Chapter 4, this concept is applied to identify additional GcvB targets. Global mRNA changes upon pulse-expression of GcvB wild-type or mutant RNAs lacking conserved part of this sRNAs are analysed as well as target-predictions refined by including a motif search for the C/A-rich GcvB target site. This reveals additional amino acid and peptide transporters as GcvB targets and in addition genes involved in amino acid biosynthesis.

Chapters 5 and 6 demonstrate the application of next generation sequencing for the identification of novel sRNAs and analysis of a whole bacterial transcriptome. In Chapter 5, deep sequencing is applied to identify *Salmonella* RNA ligands bound to the highly conserved RNA chaperone Hfq which is a key player in sRNA-mediated regulation of gene expression in bacteria. This recovers known sRNAs from *Salmonella* but also identifies novel sRNAs and mRNA targets. The bioinformatics-based work flow that was developed for the analysis and visualization for deep sequencing data is presented.

In *Helicobacter pylori* neither Hfq nor any sRNAs have been described so far. In fact, this bacterium has been referred to as an example of an organism without riboregulation (Mitarai *et al.*, 2007). A high-throughput 'RNA-seq' approach of total *H. pylori* RNA reveals in Chapter 6 that *H. pylori* in fact harbours diverse sRNA genes. In addition, a strategy for enrichment of primary transcripts in cDNA libraries is developed which allows global analyis of mRNA transcriptional start sites in the *H. pylori* genome.

Finally, the results of this study are summarized including a discussion of some future perspectives in the concluding Chapter 7. Experimental and biocomputational strategies and methods that were used or developed in this thesis are described in Chapter 8.

Previous publications

Most of the work described in Chapters 3 and 5 of this thesis has been recently published elsewhere:

- **Sharma C. M.**, Darfeuille F., Plantinga T. H., and 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 & Development**, 21(21): 2804-2817.
- Sittka A., Lucchini S., Papenfort K., **Sharma C. M.**, Rolle K., Binnewies T. T., Hinton J. C., and Vogel J. (2008) *Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq.* **PLoS Genetics**, 4(8): e1000163.
- Sittka A., **Sharma C. M.**, Rolle K., and Vogel J. (2009) *Deep sequencing of Salmonella RNA associated with heterologous Hfq proteins* in vivo *reveals small RNAs as a major target class and identifies RNA processing phenotypes.* **RNA Biology**, 6(3): 1-10.

The following manuscripts related to the work presented in Chapters 4 and 6 are in preparation or submitted:

- **Sharma C. M.**, Papenfort K., Lucchini S., Hinton J. C., and Vogel J., *GcvB RNA, a global regulator of genes involved in amino acid metabolism*, in preparation.
- **Sharma C. M.**, Hoffmann S., Darfeuille F., Findeiß S., Reignier J., Hackermüller J., Stadler P. F., and Vogel J., *The primary transcriptome of the major human pathogen, Helicobacter pylori*, in preparation.
- Hoffmann S., Otto C., Kurtz S., **Sharma C. M.**, Khaitovich P., Vogel J., Stadler P. F., and Hackermüller J., *Fast mapping of short sequences with mismatches, insertions and deletions using index structures*, submitted.

Additional work related to this thesis in a more indirect way has appeared in:

- Bouvier M., **Sharma C. M.**, Mika F., Nierhaus K. H., and Vogel J. (2008) *Small RNA binding to 5' mRNA coding region inhibits translational initiation.* **Molecular Cell**, 32(6): 827-37.
- Vogel J., and **Sharma C. M.** (2005) *How to find small noncoding RNAs in bacteria.* **Biological Chemistry**, 386(11): 1219-38.
- Mika F., **Sharma C. M.**, Bouvier M., Papenfort K., and Vogel J., *Small RNA binding to 5' mRNA coding region inhibits translational initiation*, submitted.

Contributions by others

The work described in this PhD thesis was done under supervision of Dr. Jörg Vogel in the RNA Biology group at the Max Planck Institute for Infection Biology, Berlin, Germany. Parts of the work described in this thesis which have been contributed by others or have been conducted in collaboration with others are indicated below.

1. Chapter 3: Multiple targeting of ABC transporter mRNAs by GcvB sRNA

    - Cloning of control and GcvB wild-type expression plasmids and preparation of periplasmic fractions for 2D analysis was done by Dr. Titia Plantinga (RNA Biology, MPI for Infection Biology, Berlin).
    - Identification of GcvB homologues in diverse bacteria and alignments were done in collaboration with Titia Plantinga.
    - Initial *in vitro* structure probing experiments were done jointly with Dr. Fabien Darfeuille (INSERM U869, Université Victor Segalen, Bordeaux, France).

2. Chapter 4: GcvB RNA, a global regulator of genes involved in amino acid metabolism

    - Arabinose inducible plasmids for expression of GcvB wild-type and mutant RNAs were cloned by Kai Papenfort (RNA Biology, MPI for Infection Biology, Berlin).
    - Microarray experiments were done in collaboration with Kai Papenfort.
    - Parts of the GFP fusion cloning, Western blot and FACS experiments were done with technical assistance of Franziska Seifert or by Sandy Pernitzsch during her student internship under supervision of Cynthia Sharma.

3. Chapter 5: Analysis of Hfq-bound RNAs in *Salmonella* by high-throughput sequencing

    - All wet-lab experiments described in this section (Hfq co-immunoprecipitation and Northern Blot analysis) were done by Dr. Alexandra Sittka (RNA Biology, MPI for Infection Biology, Berlin).
    - Analysis of mapping results of the deep sequencing data was done in collaboration with Alexandra Sittka.

4. Chapter 6: Deep sequencing reveals the primary transcriptome of *Helicobacter pylori*

    - For cDNA construction, *Helicobacter* growth under acid stress and infection conditions and isolation of RNA were done by Jérémy Reignier and Fabien Darfeuille (INSERM U869, Université Victor Segalen, Bordeaux, France).
    - Deep sequencing reads were aligned to the *H. pylori* genome using a new mapping method (Hoffmann *et al.*, 2009, submitted) by Dr. Steve Hoffmann (Bioinformatics Group, University of Leipzig, Germany).
    - Manual annotation of transcriptional start sites based on deep sequencing data was done in collaboration with Steve Hoffmann.

CHAPTER 2

BIOLOGICAL BACKGROUND

In addition to the major RNA components of the cell, *i. e.* tRNA, rRNA and mRNA, bacterial genomes are now known to harbour also many, perhaps several hundred, loci that encode regulatory RNAs. These RNAs are often referred to as small noncoding RNAs (ncRNAs) as they do not contain open reading frames (ORFs). Regulatory RNAs can act as RNA itself or in association with proteins in so-called ribonucleoprotein (RNP) complexes. They are involved in diverse processes, *e. g.*, transcriptional regulation, chromosome replication, RNA processing and modification, mRNA stability and translation, as well as protein degradation and translocation (Storz, 2002).

Abundant small RNA molecules other than tRNA and rRNA were first observed in *E. coli* four decades ago (Griffin, 1971; Ikemura & Dahlberg, 1973a), long before the first microRNAs (miRNAs) and short interfereing RNAs (siRNas) were discovered in eukaryotes. However, neither the genes encoding them nor their functional role was established. In the early 1980s the first plasmid-encoded antisense RNA (about 100 nucleotides in length), RNA I, was discovered and found to control plasmid-copy number in *Escherichia coli* (Stougaard *et al.*, 1981; Tomizawa *et al.*, 1981). This was followed by the identification of other antisense RNAs of mobile elements that control the life cycle or copy number of bacterial phages, transposons, and plasmids (Simons & Kleckner, 1983). The first chromosomally encoded antisense regulator, MicF RNA, was reported in 1984; it inhibits translation of the mRNA encoding the major outer membrane porin OmpF (Mizuno *et al.*, 1984). Unlike the bona fide *cis*-antisense RNAs of mobile elements MicF RNA is not transcribed from the DNA strand opposite of its target gene, *ompF*. Moreover, MicF exhibits only partial and imperfect sequence complementarity to *ompF* mRNA, yet its binding to the *ompF* mRNA near the start codon strongly inhibits the translation of this protein (Mizuno *et al.*, 1984).

Until 2001, only ten genes of such regulatory RNAs were known in *E. coli* (Wassarman *et al.*, 1999). They were discovered fortuitously using genetic screens or through direct labelling and sequencing. These RNAs included the specialized housekeeping RNAs, namely, RNase P RNA, tmRNA, and SRP RNA, which were identified as highly abundant RNA species and are involved in tRNA maturation, ribosome rescue, and protein translocation, respectively. In 2001-2002, four bioinformatics-based studies identified many new sRNA genes in *E. coli* (Argaman *et al.*, 2001; Chen *et al.*, 2002; Wassarman *et al.*, 2001; Rivas *et al.*, 2001). To date, more than 80 sRNAs are verified in *E. coli* (reviewed, *e. g.*, in Gottesman, 2004; Waters & Storz, 2009) and diverse screens have led to the identification of sRNAs in a wide range of bacteria (reviewed in Altuvia, 2007; Pichon & Felden, 2008; Vogel & Sharma, 2005).

The bacterial sRNAs range in length from ≈ 50 to ≈ 400 nt and can adopt diverse secondary structures. Figure 2.1 gives some examples of sRNAs from *E. coli* and *Salmonella*. Most of the bacterial

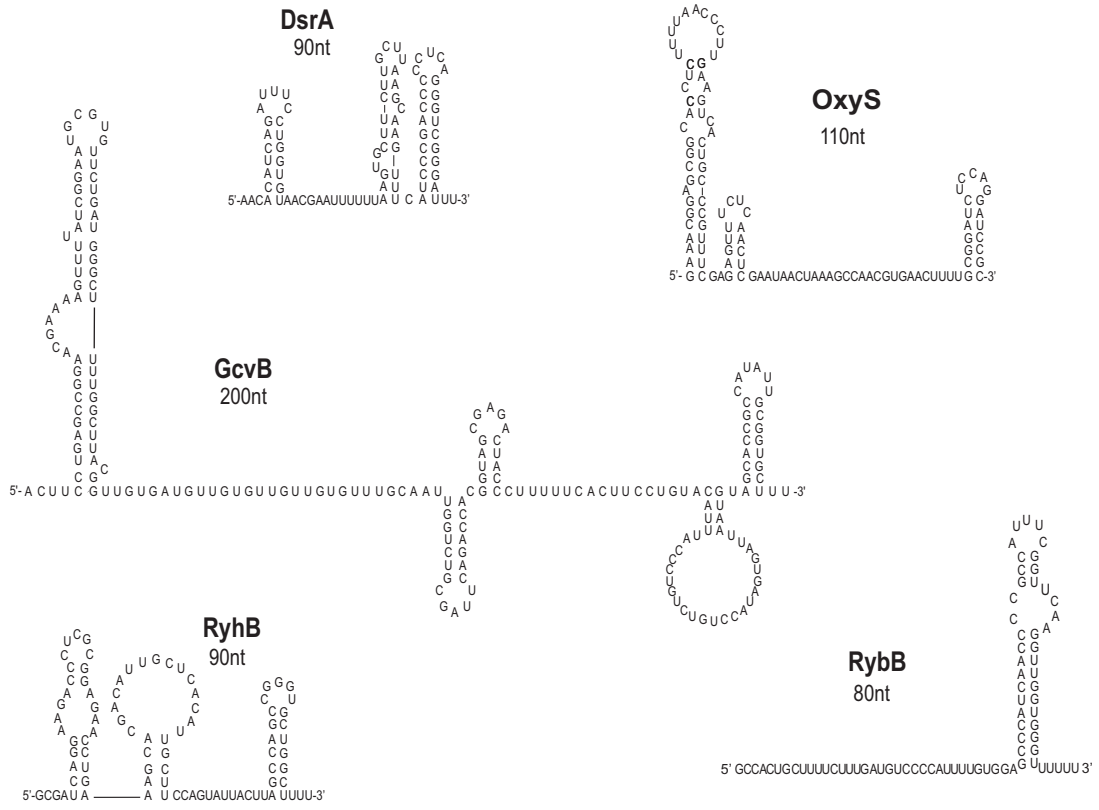


Figure 2.1: Examples of sRNAs from *Escherichia coli* and *Salmonella typhimurium*. Secondary structures of diverse *E. coli* (DsrA, OxyS, and RyhB) and *Salmonella* (GcvB, RybB) sRNAs. All of these sRNAs are conserved between *E. coli* and *Salmonella* (Hershberg *et al.*, 2003; Papenfort *et al.*, 2008). Bacterial sRNAs have different lengths and can adopt diverse secondary structures consisting of several stem-loop structures which are separated by extended single stranded regions. The secondary structures shown are based on *in vitro* structure probing experiments (DsrA: Lease & Belfort, 2000; OxyS: Altuvia *et al.*, 1997; GcvB: see Chapter 3 and Sharma *et al.*, 2007; RyhB: Geissmann & Touati, 2004; RybB: Bouvier *et al.*, 2008).

sRNAs have been shown to act as post-transcriptional regulators of gene expression in response to external stimuli and are strongly regulated under diverse growth and stress conditions (Argaman *et al.*, 2001; Wassarman *et al.*, 2001). For example, they can be induced by temperature changes (cold-shock induced DsrA: Sledjeski *et al.*, 1996), low iron (Fur-repressed RyhB: Massé & Gottesman, 2002), oxidative stress (OxyR-activated OxyS: Altuvia *et al.*, 1997), changes in glucose concentration (CRP-repressed Spot42: Møller *et al.*, 2002 and CRP-activated CyaR: De Lay & Gottesman, 2009; Johansen *et al.*, 2008; Papenfort *et al.*, 2008), elevated glucose-phosphate levels (SgrR-activated SgrS: Vanderpool & Gottesman, 2004), and outer membrane stress (σ^E-induced MicA, RybB, and VrrA: Song *et al.*, 2008 and reviewed in Vogel & Papenfort, 2006; Guillier *et al.*, 2006). This allows bacteria a rapid regulation of gene expression in response to certain stresses.

This Chapter reviews the different modes of action of bacterial sRNAs and introduces additional factors that are involved in sRNA-mediated regulation in bacteria such as the RNA-binding protein

Hfq. Furthermore, several experimental and bioinformatics-based approaches for the identification of sRNAs are discussed. Strategies to determine the mRNA targets will then pass into the recent finding that several sRNAs regulate multiple targets rather than single targets. Finally, the two model organisms, *Salmonella typhimurium* and *Helicobacter pylori*, that were used throughout this thesis are introduced.

2.1. Regulation of gene expression by regulatory RNAs in bacteria

Regulatory RNAs in bacteria constitute a heterogeneous group of molecules which act by various mechanisms to modulate gene expression in response to environmental changes. They can be encoded in *cis* or *trans* relative to their targets and activate or repress gene expression. *Cis*-encoded regulatory RNAs include riboswitches, which are part of the 5' UTRs of the mRNA that they regulate, as well as *cis*-encoded antisense RNAs encoded on plasmids or in the chromsome. Several sRNAs have been shown to bind proteins and antagonize protein activity by a sequestration mechanism. However, the majority of enterobacterial sRNAs act as antisense RNAs on *trans*-encoded mRNAs by imperfect base-pairing and thereby inhibit translation and/or lead to mRNA degradation (Majdalani *et al.*, 2005; Storz *et al.*, 2005). Therefore, the latter class is often regarded as functionally analogous to eukaryotic miRNAs.

2.1.1. Translational repression

The above mentioned MicF RNA was only the first example of an emerging class of *trans*-encoded antisense RNAs (Mizuno *et al.*, 1984). For example, two other major outer membrane proteins, OmpA and OmpC, have also been shown to be regulated at the translational level by their cognate sRNAs, MicA and MicC (Chen *et al.*, 2004; Udekwu *et al.*, 2005). The canonical model of these sRNAs is to mask the ribosome binding sites (RBS) of their target via imperfect sequence complementarity and, thus, inhibit ribosome entry (Storz *et al.*, 2004). Specifically, the *E. coli* MicA, MicC, OxyS, and Spot42 sRNAs were shown to directly interfere with 30S ribosome binding of their target mRNAs, *ompA*, *ompC*, *fhlA*, and *galK*, respectively, by *in vitro* toeprinting experiments (see Figure 2.2A and Argaman & Altuvia, 2000; Chen *et al.*, 2004; Møller *et al.*, 2002; Udekwu *et al.*, 2005). Other well-characterized *E. coli* sRNA-target pairs that are likely to use the same mechanism include DsrA-*hns*, MicF-*ompF*, RyhB-*sodB*, and SgrS-*ptsG* (reviewed in Wagner & Darfeuille, 2006). Furthermore, sRNAs from other enterobacteria have been shown to act by the same mechanism, *e. g.*, *ompA* targeting by VrrA RNA in *Vibrio cholerae* (Song *et al.*, 2008). In addition, also sRNAs from Gram-positive bacteria, such as RNAIII from *Staphylococcus aureus* and SR1 from *Bacillus subtilis*, have been shown to inhibit translation by interfering with ribosome binding (Boisset *et al.*, 2007; Geisinger *et al.*, 2006; Heidrich *et al.*, 2007; Huntzinger *et al.*, 2005).

Since the half-life of bacterial mRNAs is strongly affected by the association with ribosomes (Deana & Belasco, 2005), translation inhibition will promote the decay of the repressed target, *e. g.*, by accelerating RNase E-mediated mRNA turnover (Massé *et al.*, 2003; Morita *et al.*, 2005). Besides RNase E, the bacterial Sm-like protein Hfq has been identified as a key player in this type of translational silencing (see below, Section 2.2.1). Hfq binds all of the aforementioned sRNAs with high affinity and is most often required for both their intracellular stability and their interaction with

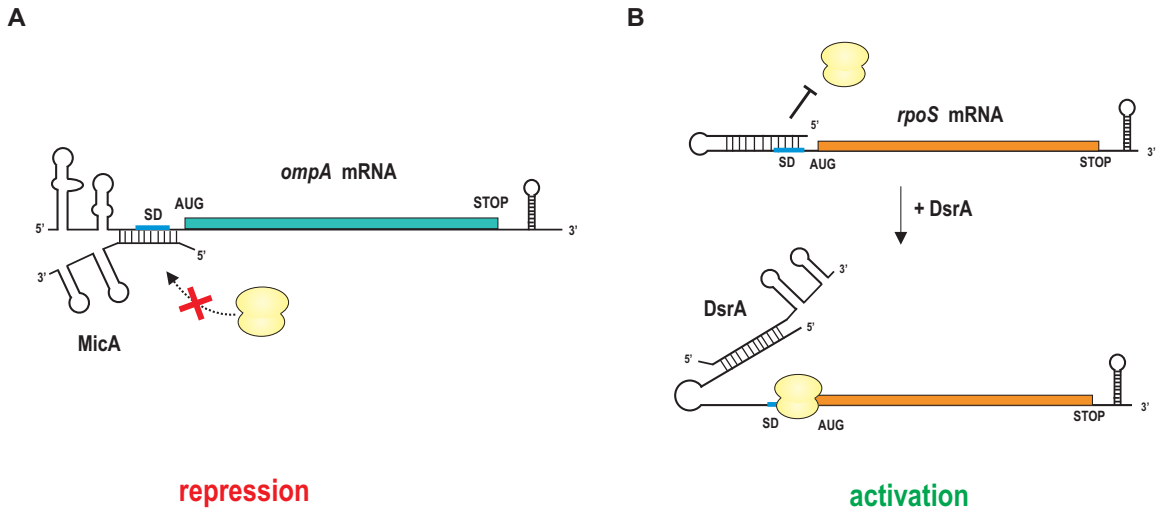


Figure 2.2: Translational repression and activation mediated by bacterial sRNAs. (A) sRNA-mediated translational repression. Binding of MicA RNA to the 5' UTR of *ompA* mRNA masks the Shine-Dalgarno sequence and thereby inhibits translation initiation (Udekwu *et al.*, 2005). **(B)** sRNA-mediated translational activation. Under non-activating conditions, the *rpoS* 5' UTR folds into an inhibitory stem-loop structure including nucleotides closely upstream of the AUG start codon which occludes the SD sequence and thereby inhibits translation (Brown and Elliot, 1997). Interaction of DsrA RNA with *rpoS* mRNA leads to translational activation by dissolving the fold-back structure of *rpoS* mRNA in which the ribosome binding site is masked (Majdalani *et al.*, 1998). Secondary-structures of sRNAs and mRNAs are shown schematically. The SD sequences are indicated in blue, and ORFs in light blue (*ompA*) and orange (*rpoS*).

target mRNAs (Aiba, 2007; Urban & Vogel, 2007; Zhang *et al.*, 2003, and references therein). In addition, Hfq and some sRNAs were found to form ribonucleoprotein complexes with RNase E to mediate target mRNA destabilization (Morita *et al.*, 2005).

Translational repression by bacterial sRNAs is not limited to the 5' UTRs of monocistronic mRNAs as also intergenic regions of polycistronic messenger RNAs can be targeted by sRNAs. For example, in *E. coli*, the *galETKM* operon, which encodes components involved in galactose metabolism, is targeted by Spot42 sRNA at internal sequences of the polycistronic mRNA. Upon induction of the sRNA in response to high glucose levels, Spot42 occludes the RBS of the *galK* cistron and inhibits its translation without affecting the upstream *galET* cistrons (Møller *et al.*, 2002). Therefore, this case constitutes the prototype of sRNA-mediated discoordinate operon-expression.

Several sRNAs have recently been shown to repress multiple mRNAs (see Section 2.5). These include, *e. g.*, RybB, OmrA and OmrB, which regulate expression of several outer membrane proteins (Guillier & Gottesman, 2006; Papenfort *et al.*, 2006) or RyhB RNA which regulates multiple genes involved in iron metabolism (Massé *et al.*, 2005). Most of the characterized antisense RNAs in bacteria inhibit translation by competing with ribosomes for translation initiation regions (TIR) on nascent mRNA. However, recent studies show that translational repression can also be achieved by binding of an sRNA far upstream or downstream of the translation initiation site. One of these RNAs is GcvB RNA which will be discussed in Chapters 3 and 4 of this thesis. Another example is IstR-1 RNA in *E. coli* (Vogel *et al.*, 2004) which prevents translation of the SOS-induced toxin

TisB by binding ≈100 nt upstream of the *tisB* RBS under non-SOS conditions. Instead of targeting the SD sequence of *tisB*, IstR-1 competes with ribosomes by base pairing with a ribosome loading or "standby" site which is required for initiation of translation at the highly structured *tisB* TIR. Recently, *Salmonella* RybB RNA was found to repress *ompN* mRNA translation by pairing with the 5' coding region (Bouvier *et al.*, 2008). Systematic analysis of antisense interference with 30S binding revealed that sequestering sequences within the mRNA down to the fifth codon allows sRNAs to act as translational repressors.

2.1.2. Translational activation

Bacterial sRNAs can not only repress mRNAs but, in fact, can also upregulate gene-expression by an anti-antisense mechanism. Besides the best known example, DsrA, several other sRNAs have been shown to activate gene expression by disrupting an inhibitory secondary structure which sequesters the ribosome binding site and thereby prevents translation in the absense of the sRNA (reviewed in Waters & Storz, 2009). Therefore, activation of translation seems to be more widespread than previously anticipated. In *E. coli*, DsrA RNA was found to be a translational activator of the major stress and stationary phase sigma factor, RpoS, at low growth temperatures (Majdalani *et al.*, 1998; Sledjeski *et al.*, 1996). The *rpoS* mRNA contains an extraordinarily long 5' UTR (≈600 nt), which can fold into a translationally-inactive structure by masking the RBS. Upon base-pairing with DsrA RNA the stem-loop structure is opened and the RBS gets accessible for ribosome binding (see Fig. 2.2B). A search for additional sRNAs that modulate RpoS expression resulted in the identification of the 105 nt RprA RNA (Majdalani *et al.*, 2001). Even though both sRNAs, DsrA and RprA, pair with the same region ≈100 nt upstream of the RBS and disrupt hairpin formation (Majdalani *et al.*, 2001, 1998), the two sRNAs act under different conditions on *rpoS* mRNA. While DsrA is induced at low temperatures, RprA expression peaks upon cell surface stress (Majdalani *et al.*, 2002; Repoila & Gottesman, 2001).

Similar to repressing sRNAs, activating sRNA can also mediate discoordinate operon expression. Two small RNAs from *E. coli*, GlmY and GlmZ, have recently been shown to mediate discoordinate operon expression of the *glmUS* mRNA in which the downstream gene *glmS*, an essential enzyme in amino-sugar metabolism, is activated at the post-transcriptional level (Kalamorz *et al.*, 2007; Urban *et al.*, 2007). Although the two sRNAs are highly similar in sequence and structure, they act in a hierarchical manner. GlmZ, together with the RNA chaperone Hfq, directly activates *glmS* mRNA translation by disruption of an inhibitory mRNA structure similar to DsrA-*rpoS* (Reichenbach *et al.*, 2008; Urban & Vogel, 2008). In contrast, GlmY acts upstream of GlmZ and positively regulates *glmS* by stabilizing GlmZ RNA. The current model assumes that this is achieved by titrating an RNA-processing protein, YhbJ, which processes GlmZ and abolishes its ability to activate *glmS* translation. Thus, GlmY competes with GlmZ for binding to the YhbJ protein and activates *glmS* expression indirectly by increasing levels of active GlmZ RNA.

Activation of gene expression can also be mediated by *cis*-encoded RNAs. For example, in *E. coli*, the stationary phase-induced GadY RNA is transcribed from the opposite strand to its target mRNA, *gadX*, a transcriptional regulator of the acid response (Opdyke *et al.*, 2004; Tramonti *et al.*, 2008). Base pairing between GadY and the *gadX* 3´ UTR leads to cleavage of the *gadXW* polycistronic mRNA, stabilization of *gadX* mRNA and, in turn, to accumulation of the GadX protein.

2.1.3. sRNA mediated sequestration of protein activity

Messenger RNAs are not the only targets of sRNAs. Several *E. coli* sRNAs, and their homologues in other bacteria, interact with cellular proteins to modulate their activities. They antagonize the activity of their target proteins by mimicking the structures of other RNAs or DNA and thereby affect transcription, translation or processing of other RNAs. The best known examples are 6S RNA which titrates RNA polymerase and the CsrB/CsrC RNAs which antagonize the CsrA protein. These RNAs are discussed below in more detail. Furthermore, a plasmid-encoded RNA was found to mediate stability of ColE1 plasmids by modulating tryptophanase activity in *E. coli* (Chant & Summers, 2007). Specifically, Rcd RNA is transcribed from ColE1-plasmid dimers which can occur during replication. Binding of Rcd RNA to tryptophanase leads to a higher affinity of the enzyme's substrate trypotphan and in turn increased production of indole. This causes a cell division arrest and allows more time to resolve plasmid dimers and to maintain plasmids.

2.1.3.1. 6S RNA

6S RNA, one of the most abundant and conserved RNAs in bacteria, was first detected by *in vivo* RNA labelling experiments (Hindley, 1967) and subsequently sequenced by enzymatic digestion (Brownlee, 1971), but its gene was identified much later (Hsu *et al.*, 1985). It is co-transcribed with the downstream *ygfA* gene and processed from this dicistronic transcript (Hsu *et al.*, 1985; Kim & Lee, 2004). 6S RNA accumulates in stationary phase (Barrick *et al.*, 2005) and folds into a rod-like structure that is required for binding to the σ^{70}-containing RNA-polymerase holoenzyme (Trotochaud & Wassarman, 2005; Wassarman & Storz, 2000). It is thought to mimic the DNA template of the open promoter complex of RNA polymerase (see Figure 2.3A; reviewed in Wassarman, 2007). Gel-shift experiments showed that 6S RNA interacts with σ^{70}-RNA polymerase but is not associated with σ^{S}-RNA polymerase (Trotochaud & Wassarman, 2005). This induces a change in the holoenzyme's promoter recognition specificity and inhibits transcription from certain σ^{70} promoters, whereas transcription from some σ^{S} promoters is increased (Trotochaud & Wassarman, 2004, 2005; Wassarman & Storz, 2000). Recently, it has been shown that 6S RNA not only mimics an open promoter complex but rather can serve as a template for transcription of 14-20 nt product RNAs, so-called pRNAs (Gildehaus *et al.*, 2007; Wassarman & Saecker, 2006). It was suggested that synthesis of the pRNAs is required for release of the RNA polymerase from 6S RNA during outgrowth from stationary phase or increased NTP concentrations; whether the pRNAs themselves have a function remains elusive.

Until recently, homologous sequences of *E. coli* 6S RNA were only known in the γ-subdivision of proteobacteria. However, based on a covariance model of 6S RNA structure, multiple 6S RNA homologues could recently be identified in more than 100 bacterial species of diverse eubacterial lineages (Barrick *et al.*, 2005). During this comprehensive screen, a number of abundant *Bacillus* and *Synechococcus* RNAs of previously unknown function turned out to be 6S RNA homologues (Ando *et al.*, 2002; Suzuma *et al.*, 2002; Watanabe *et al.*, 1997). Biochemical analysis by others independently confirmed these *Bacillus* RNAs as 6S RNA homologues (Trotochaud & Wassarman, 2005). Strikingly, although the *E. coli* 6S RNA and each of the two *Bacillus* 6S RNAs share only $\approx 46\,\%$ similarity of the primary sequence, key secondary structure elements and functionally important residues are conserved. An extended covariance model which includes candidates with

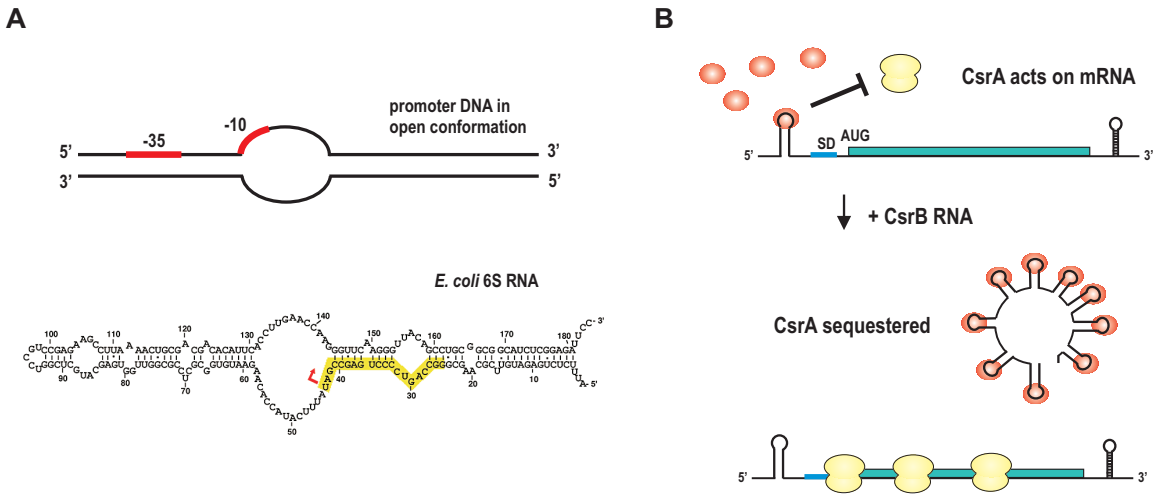


Figure 2.3: Sequestration of protein activity by bacterial sRNAs. (A) DNA in open pomoter conformation and 6S RNA. Figure adapted from Wassarman (2007). **(B)** The CsrA protein (red circles) binds to hairpins in mRNAs and thereby in inhibits translation. When CsrB RNA is expressed, multiple stem-loops in the sRNA bind to CsrA proteins and, thus, sequester protein activity. This allows translation of otherwise repressed mRNAs.

slightly deviating structures predicted hundreds of additional 6S RNA sequences in diverse microbial genomes (Barrick *et al.*, 2005). Furthermore, an RNomics-based approach lead to the identification of an 6S homologue in the hyperthermophilic bacterium *Aquifex aeolicus* (see Section 2.3.5 and Willkomm *et al.*, 2005).

2.1.3.2. CsrB, CsrC

In *E. coli*, two small RNAs, CsrB and CsrC, act as antagonists of the carbon storage regulator protein, CsrA (reviewed in Babitzke & Romeo, 2007). CsrA modulates (usually inhibits) translation by binding to GGA motifs in the 5' UTRs of certain target mRNAs, including some that are involved in glycogen biosynthesis (see Figure 2.3B). The two conserved RNAs form a regulatory feedback loop with CsrA protein and, thereby, tightly control the active pool of the protein (Romeo, 1998; Weilbacher *et al.*, 2003). Instead of directly acting upon mRNAs, CsrB and CsrC commonly sequester the abundant RNA-binding protein, CsrA. Both RNAs contain multiple hairpin structures with GGA sequence motifs in the loops which serve as CsrA binding sites (see Figure 2.3B and Liu *et al.*, 1997; Weilbacher *et al.*, 2003). Thus, upon increase of CsrB and CsrC levels the sRNAs effectively sequester the CsrA protein which, in turn, directly modulates mRNA translation (Babitzke & Romeo, 2007). Many bacterial species contain the global regulator CsrA as well as homologues of its two regulatory sRNAs, CsrB and CsrC. These include, for example, CsrB, CsrC and CsrD RNA involved in quorum sensing in *Vibrio cholerae* (Lenz *et al.*, 2005), PrrB RNA, the functional CsrB homologue in the biocontrol strain *Pseudomonas fluorescens* F113 (Aarons *et al.*, 2000), the RsmB RNAs in diverse *Erwinia* species (Liu *et al.*, 1998; Ma *et al.*, 2001), and the RsmY and RsmZ RNAs from *Pseudomonas* (Heeb *et al.*, 2002; Valverde *et al.*, 2003). Although the described CsrB homologues from a variety of organisms have little similarity at the primary sequence level, these

RNAs do share significant similarity at the secondary structure level and in the frequency of occurrence of GGA repeats, which are required for CsrA/RsmA recognition (see Romeo, 1998; Valverde *et al.*, 2004).

2.1.4. Riboswitches and RNA thermosensors

Riboswitches and RNA thermosensors are *cis*-encoded regulatory elements which are usually found in the 5' UTRs of the mRNAs that they regulate in response to environmetal signals, for example stalled ribosomes, uncharged tRNAs, temperature, or small molecule ligands (reviewed in Grundy & Henkin, 2006). Riboswitches are built of an aptamer region which can bind a ligand and an expression platform which regulates gene expression through alternative secondary structures. This can result in transcription attenuation by formation of a transcriptional terminator or translational repression by forming a structure where the RBS is occluded (reviewed in Mandal & Breaker, 2004; Mironov *et al.*, 2002). Generally, the aptamers sense and bind metabolite ligands, such as purines, *S*-adenosyl-methionine, flavin mononucleotide, or lysine. Recently, an RNA sensor of Mg^{2+} was identified in the 5' UTR of *mgtA* mRNA, which encodes an Mg^{2+} transporter (Cromie *et al.*, 2006). In addition to ligand-binding riboswitches, RNA structures can also sense temperature changes which lead to refolding of the 5' UTR and thereby control gene expression. In *Listeria monocytogenes*, a thermosensor was identified in the 5' UTR of *pfrA*, encoding an important transcriptional activator of virulence genes (Johansson *et al.*, 2002). At 30°C, this sensor forms an inhibitory structure which sequesters the RBS; however, a higher temperature (37°C) leads to an opening of the stem-loop structure which, in turn, allows translation of the mRNA. In *Salmonella* similar RNA thermometers were discovered in the 5' UTRs of *ibpA* and *agsA* mRNAs (Waldminghaus *et al.*, 2005, 2007).

Most of the riboswitches in Gram-positive bacteria act by transcriptional attenuation, whereas riboswitches in Gram-negative bacteria mainly repress translation (reviewed in Mironov *et al.*, 2002). Furthermore, Gram-positive bacteria rely more on *cis*-acting riboswitches, contrary to Gram-negative bacteria for which more *trans*-acting RNAs are known. For example, the *glmS* gene is controlled by a self-cleaving riboswitch in Gram-positive bacteria (Collins *et al.*, 2007; Winkler *et al.*, 2004), whereas in *E. coli* the two above mentioned *trans*-encoded sRNAs, GlmY and GlmZ, regulate *glmS* expression.

2.1.5. *Cis*-endoded sRNAs

As mentioned above, the first bacterial antisense RNAs were identified on plasmids (reviewed in Wagner *et al.*, 2002). Many of these *cis*-encoded plasmid sRNAs modulate the synthesis of replication proteins, *e. g.*, by inhibition of replication primer formation by RNAI and, thus, control the plasmid copy number (see Brantl, 2007). The *cis*-encoded RNAs are transcribed as distinct RNAs from the opposite strand of their targets, and initial interactions to their targets are often mediated by loop-loop interactions and followed by extension of the duplex (reviewed in Brantl, 2007). Other *cis*-encoded sRNAs from plasmids repress the synthesis of toxic proteins and act as plasmid-addiction molecules or post-segregational killing systems (Gerdes *et al.*, 1997). The best known toxin-antitoxin pair is the *hok/sok* system of plasmid R1 (Gerdes *et al.*, 1990). Toxin-antitoxin systems that are homologous to the Hok/Sok system have been identified not only on plasmids but also

in the *E. coli* chromosome (for a review see Fozo *et al.*, 2008a). Further chromosomally encoded *cis*-acting antisense RNAs have been found in cryptic prophages, such as RatA RNA in *B. subtilis* (Silvaggi *et al.*, 2005). Recently, several small toxic proteins that are present in multiple copies were identified in the *E. coli* chromosome and were found to be repressed by *cis*-encoded antisense RNAs (Fozo *et al.*, 2008b). Other examples for *cis*-encoded antisense RNAs are SymR RNA in *E. coli* which is transcribed opposite to the 5' end of the SOS-induced toxin SymE and thereby inhibits translation of its target (Kawano *et al.*, 2007), or the iron-stress repressed IsrR RNA which is located within the *isiA* coding region and was described to control expression of this photosystem associated protein in *Synechocystis* (Dühring *et al.*, 2006).

Similar to *trans*-encoded sRNAs, *cis*-encoded antisense RNAs can also mediate discoordinate operon expression. These include GadY RNA which stabilizes *gadX* mRNA by cleavage of the *gadXW* polycistronic transcript (see Section 2.1.2). In *Vibrio anguillarum*, a plasmid-encoded antisense RNA mediates processing of a polycistronic mRNA. However, in this case sRNA binding leads to transcription termination after the *fatA* gene of the *fatDCBAangRT* operon and, hence, reduces expression of the downstream *angRT* genes (Stork *et al.*, 2007).

2.1.6. sRNAs with dual functions

Although sRNAs were for a long time considered to be non-coding RNAs, several sRNAs are now known to be bifunctional, *i. e.* they contain an open reading frame and in addition act as an antisense RNA. Thus, regulatory RNAs do not necessarily have to be noncoding RNAs. One example is the highly structured RNAIII of *Staphylococcus aureus* which encodes the 26 amino-acid peptide δ-hemolysin in its 5' part and modulates the expression of two other virulence genes, *e. g.*, *spa* and *hla*, through base-pairing of its noncoding regions with the mRNAs (Boisset *et al.*, 2007; Huntzinger *et al.*, 2005; Novick *et al.*, 1993). In *E. coli*, the 227-nt long SgrS RNA, which is expressed during glucose-phosphate stress, was shown to contain an 43 amino acid ORF, SgrT, upstream of the nucleotides that are involved in the antisense function of this sRNA (Wadler & Vanderpool, 2007). SgrS was previously shown to block translation of the *ptsG* mRNA encoding a sugar-phosphate transporter by means of a base-pairing dependent mechanism requiring the RNA chaperone Hfq (Kawamoto *et al.*, 2006; Vanderpool & Gottesman, 2004). It was suggested that SgrT reinforces the regulation by SgrS by independently downregulating glucose uptake by directly or indirectly inhibiting the PtsG transporter (Wadler & Vanderpool, 2007).

2.2. Additional factors involved in gene regulation by bacterial sRNAs

Small regulatory RNAs in bacteria either pair with mRNA targets or modify protein activities. Moreover, they often act in ribonucleoprotein complexes. Proteins that interact with the sRNAs can possess catalytic activity, induce conformational changes of the sRNA, or be sequestered by the sRNA to prevent the action of the protein. The current knowledge of the various proteins that interact with RNA regulators and the physiological implications of sRNA-protein complexes in DNA, RNA and protein metabolism, as well as in RNA and protein quality control are reviewed in Pichon & Felden (2007). Here, the role of the key players Hfq and several RNases in sRNA-mediated regulation in bacteria is introduced.

2.2.1. The Sm-like protein Hfq

In bacteria, the majority of the sRNAs base-pair with target mRNAs to regulate their translation and/or decay and these regulatory events commonly require the bacterial Sm (Smith antigen)-like protein, Hfq. Hfq is a 102 amino acid protein that was first identified in *Escherichia coli* as a host factor required for phage Qβ RNA replication $\approx$ 40 years ago (Franze de Fernandez *et al.*, 1968). Hfq is now known to have important physiological roles in numerous model bacteria (Valentin-Hansen *et al.*, 2004). Almost half of all sequenced Gram-negative and Gram-positive species, and at least one archaeon, encode an Hfq homologue (Nielsen *et al.*, 2007; Sun *et al.*, 2002). In *E. coli*, it is a highly abundant protein with estimated 50,000 to 60,000 copies ($\approx$ 10,000 hexamers) per cell, the majority being associated with ribosomes and a minor fraction with the nucleoid (Azam *et al.*, 1999; Franze de Fernandez *et al.*, 1972; Kajitani *et al.*, 1994). The crystal structures of Hfq proteins from *Staphylococcus aureus* (Schumacher *et al.*, 2002), *E. coli* (Sauter *et al.*, 2003), and *Pseudomonas aeruginosa* (Nikulin *et al.*, 2005) revealed that Hfq is an Sm protein. The eukaryotic and archaeal Sm and Sm-like (Lsm) proteins form heteroheptameric rings and are central components of RNP complexes involved in diverse aspects of RNA metabolism including splicing and mRNA decay (reviewed in Wilusz & Wilusz, 2005). In *E. coli* and related bacteria, Hfq forms homohexameric rings and preferentially binds A/U-rich single-stranded regions preceded or followed by a stem-loop structure (reviewed in Brennan & Link, 2007). It has at least two major RNA binding sites, a proximal site, first identified in the crystal structure of *S. aureus* Hfq bound to the hepta-oligoribonucleotide, AU_5 (Schumacher *et al.*, 2002), and a distal site which binds poly(A) tails (Brennan & Link, 2007).

The importance of Hfq for sRNA-mediated regulation was first evident in studies on OxyS RNA (Zhang *et al.*, 1998). By now it is known that Hfq interacts with most of the regulatory sRNAs as well as diverse mRNAs (Sittka *et al.*, 2008; Zhang *et al.*, 2003) and is required for the intracellular stability of many regulatory sRNAs (Valentin-Hansen *et al.*, 2004). Hfq turned out to have an RNA chaperone activity as structure changes have been observed for some sRNAs (*e. g.* OxyS and Spot42) and mRNAs (*e. g. sodB* and *ompA*) in structure probing experiments (Geissmann & Touati, 2004; Moll *et al.*, 2003). Furthermore, it was found that Hfq possesses ATPase acitivity which might be related to its chaperone function (Sukhodolets & Garges, 2003). Recently, a plausible ATP-binding site in Hfq was identified by biochemical and genetic techniques, and it was suggested that ATP binding by the Hfq-RNA complex results in its significant destabilization (Arluison *et al.*, 2007b).

The proposed role of Hfq in sRNA-mediated posttranscriptional regulation is to facilitate the generally short and imperfect base-pairing between sRNAs and their mRNA targets (Kawamoto *et al.*, 2006; Lease & Woodson, 2004; Mikulecky *et al.*, 2004; Møller *et al.*, 2002; Zhang *et al.*, 2002, 2003). Recent fluorescence resonance energy transfer (FRET) studies showed that Hfq enhances RNA annealing and promotes strand exchange by binding rapidly to both DsrA and *rpoS* mRNA (Arluison *et al.*, 2007a; Rajkowitsch & Schroeder, 2007). Although it is known that Hfq binds to sRNAs and several mRNAs, the mechanisms by which Hfq promotes interactions remain unclear. Due to its chaperone acitivity it could promote base-pairing by opening the regions of pairing or facilitate base-pairing by increasing the local concentrations of the interacting RNAs (Storz *et al.*, 2004). In addition, it is unclear whether one Hfq hexamer binds the sRNA and mRNA simultane-

ously or if one hexamer binds the sRNA and a second hexamer the mRNA and the two hexamers could be brought together via interactions between the hydrophobic backs (Schumacher *et al.*, 2002; Storz *et al.*, 2004).

Contrary to most sRNAs from Gram-negative bacteria which are dependent on Hfq, sRNA-mediated regulation in some Gram-positive bacteria has been shown to function also without Hfq. For example, although *S. aureus* Hfq binds RNAIII, it has no detectable effect on RNAIII-target mRNA complex formation, as the RNAs interact rapidly and do not require Hfq for their annealing. Furthermore, the *hfq* gene is transcribed very weakly in multiple strains of this bacterium and an *hfq* deletion strain has no detectable phenotypic effects on virulence (Bohn *et al.*, 2007; Geisinger *et al.*, 2006). Likewise, the small RNA SR1 and its target *ahrC* were shown to bind Hfq; however, Hfq was not required to promote *ahrC*/SR1 complex formation but to enable the translation of *ahrC* mRNA (Heidrich *et al.*, 2007).

In addition to its role in regulation of gene expression in a concerted manner with sRNAs, Hfq can also act alone as a translational repressor of mRNA (Urban & Vogel, 2008; Vytvytska *et al.*, 2000). Moreover, it can modulate the decay of some mRNAs, *e. g.*, by binding to their poly(A) tails and stimulating polyadenylation or protecting messages from polynucleotide phosohorylase (PNP), RNase II, and RNase E by sequestering binding sites for these RNases (Folichon *et al.*, 2005, 2003; Hajnsdorf & Régnier, 2000; Mohanty *et al.*, 2004). Hfq has also been found to autoregulate its own expression at the translational level, and roles of Hfq in tRNA biogenesis have recently been described (Lee & Feig, 2008; Scheibe *et al.*, 2007; Vecerek *et al.*, 2005). Besides binding RNAs, Hfq interacts with several components of the ribosome, degradosome or other cellular machines that are involved in RNA metabolism, *e. g.*, RNase E, PNP, poly(A) polymerase I (PAP I) (see Section 2.2.2, or RNA polymerase in a ribosomal protein S1-dependent manner (Sukhodolets & Garges, 2003).

The pleiotropy of an *hfq* deletion mutant was first apparent from the multiple stress response-related phenotypes in *E. coli*, including decreased growth rates, increased sensitivity to ultraviolet light, mutagens and oxidants, and increased cell length (Tsui *et al.*, 1994). Moreover, an *hfq* mutation changes expression of more than 50 proteins which is partly reflected by the reduced efficiency of translation of *rpoS* mRNA, encoding the major stress sigma factor, σ^S (Brown & Elliott, 1996; Muffler *et al.*, 1996). However, Hfq clearly influences bacterial physiology in a much broader fashion, including the σ^S-independent control of virulence factors in diverse pathogenic bacteria (Christiansen *et al.*, 2006; Ding *et al.*, 2004; Fantappiè *et al.*, 2009; Meibom *et al.*, 2009; Nakao *et al.*, 1995; Robertson & Roop, 1999; Sharma & Payne, 2006; Sonnleitner *et al.*, 2003). A strong impact on virulence as well as other stress-related phenotypes were also observed for a *Salmonella hfq* deletion strain (see Section 2.6.1.1 and Sittka *et al.*, 2007).

2.2.2. Ribonucleases

Besides the RNA-chaperone Hfq, several ribonucleases (RNases) are involved in sRNA-mediated regulation in bacteria and have been shown to influence the processing and turnover of these molecules (Viegas *et al.*, 2007). Since RNases are key modulators of RNA decay, the identification of the RNases that contribute to the decay of individual sRNAs is essential for a more general understanding of sRNA turnover *in vivo* (reviewed in Viegas & Arraiano, 2008). For example, poly(A)

polymerase I modulates RNA stability by adding poly(A) tails to the 3' end of RNAs which promote efficient exonucleolytic degradation (Hajnsdorf *et al.*, 1995; O'Hara *et al.*, 1995; Sarkar, 1997; Xu & Cohen, 1995). PAP I can be a main factor involved in mRNA decay but was also shown to affect stability of sRNAs such as GlmY (Urban & Vogel, 2008). In addition, Hfq was co-purified in an sRNA-independent manner with PNP and PAP I, two components of the degradosome (Mohanty *et al.*, 2004).

One of the major endoribonucleases in *E. coli* is RNase E which is a single-strand specific endoribonuclease with a main role in mRNA decay but being also involved in the processing of ribosomal and transfer RNAs (Arraiano & Maquat, 2003; Carpousis, 2002; Régnier & Arraiano, 2000). RNase E is also one of the main enzymes forming the degradosome, a multiprotein complex involved in the decay of many RNAs (Carpousis *et al.*, 1994, 1999). It cleaves single-stranded regions of structured RNAs and has a preference for 5' monophosphate termini and AU-rich sequences of RNAs (Ow *et al.*, 2003). RNase E is not ubiquitous in bacteria but functional homologues, RNase J1 and J2, were identified in *Bacillus subtilis* (Even *et al.*, 2005). Besides a role in decay of single RNAs, RNase E is important for coupled sRNA-mRNA degradation (Afonyushkin *et al.*, 2005; Massé *et al.*, 2003; Morita *et al.*, 2005). Specifically, RNase E was found to co-purify together with two sRNAs, SgrS and RyhB, and Hfq and leads to degradation of the mRNA targets *ptsG* and *sodB* (Morita *et al.*, 2005, 2006). In contrast to the degradosome, these specialized ribonucleoprotein complexes do not contain PNP, enolase, and the RhlB RNA helicase. Moreover, the Hfq binding site of RNase E was identified in the C-terminal scaffold domain which is also required for RyhB-mediated degradation of *sodB* mRNA (Massé *et al.*, 2003). It was proposed that Hfq and the degradosome components, except enolase, compete for RNase E binding at the C-terminal scaffold domain. Overall, the RNase E-Hfq-sRNA RNP complex leads to translational repression and rapid target mRNA degradation. However, Hfq binding in the absence of RNase E and RNA-RNA interaction itself are sufficient to mediate translational repression, destabilization, and destruction of the target mRNA (Maki *et al.*, 2008; Morita *et al.*, 2006).

Another RNase which is involved in post-transcriptional regulation by bacterial sRNAs is the double-strand specific RNase III. This ubiquitous RNase plays multiple roles in the processing of rRNA and mRNA (Nicholson, 1999) and can also affect the decay of some mRNAs (Régnier & Grunberg-Manago, 1990; Santos *et al.*, 1997). Interestingly, maturation of eukaryotic siRNAs and miRNAs is mediated by members of the RNase III family, namely, the double-strand specific enzymes Dicer and Drosha (Bernstein *et al.*, 2001). In *E. coli*, an antisense interaction between the SOS-induced small RNA IstR-1 and its target *tisAB* was found to entail RNase III-dependent cleavage and thereby inactivates the mRNA for translation (Vogel *et al.*, 2004). Furthermore, the decay of RyhB *in vivo* was shown to be mainly dependent on RNase III in contrast to the RNase E-dependent turnover of its target *sodB* mRNA. Cleavage of RyhB by RNase III *in vitro* is facilitated upon base-pairing with the *sodB* 5'-UTR (Afonyushkin *et al.*, 2005). Moreover, RNase III is important for regulation of several virulence factors by RNAIII in *Staphylococcus aureus* (Boisset *et al.*, 2007; Huntzinger *et al.*, 2005). In this case, coordinated action of RNase III is essential to degrade the mRNA and irreversibly arrest translation *in vivo*.

2.3. Approaches for the identification of bacterial sRNAs

Due to the diverse functions that are associated with the different structures and features of sRNAs it is difficult to find a universal method that would allow to detect all classes of sRNAs. The detection of ncRNAs by biochemical and genetic methods is quite difficult because sRNAs lack characteristic features like a poly(A)-tail and have mostly only a small size making them a poor target for mutational screens. Computational identification is also difficult; conventional protein-gene finding programs search for features like open reading frames and exon/intron boundaries, which are irrelevant to ncRNAs. The first sRNAs were identified fortuitously using genetic screens or through radiolabelling of total RNA and isolation from gels (Wassarman *et al.*, 1999). Four initial systematic screens in *E. coli* in 2001 and 2002 revealed more than 50 new sRNAs and generated an even longer list of sRNA candidate loci (Argaman *et al.*, 2001; Chen *et al.*, 2002; Rivas *et al.*, 2001; Wassarman *et al.*, 2001). Up to now, several experimental and bioinformatics-based approaches have been developed and have led to the identification of sRNAs in a wide range of bacteria, including several pathogens (see Table 2.1). Some of these approaches which will be explained in the following and include, *e. g.*, bioinformatics-based prediction strategies, shotgun-cloning of RNA (RNomics), hybridization on microarrays or co-immunoprecipitation with Hfq. Furthermore, the different identification strategies of sRNAs in diverse bacteria are summarized in several reviews (Altuvia, 2007; Hüttenhofer & Vogel, 2006; Livny & Waldor, 2007; Pichon & Felden, 2008; Vogel & Sharma, 2005).

2.3.1. Direct Labelling

The first bacterial small RNAs, other than tRNAs and 5S rRNA, were found by gel-fractionation of metabolically labelled *E. coli* total RNA (Griffin, 1971; Hindley, 1967; Ikemura & Dahlberg, 1973a,b). These studies used radio-labelled orthophosphate ($^{32}PO_4^{3-}$) which is readily taken up by growing bacteria and incorporated into nucleic acids. Following such treatment, total cellular RNA was isolated, separated by gel fractionation, and selected bands or spots sequenced by digestion with nucleases (fingerprinting) after excision from the gel. This approach identified the housekeeping RNAs, RNase P RNA, tmRNA, and SRP RNA, as well as other abundant regulatory RNAs such as 6S RNA, and Spot 42 RNA. Instead of tedious nuclease-fingerprinting assays, the sequence of isolated RNAs can also be determined by rapid cDNA cloning and Sanger sequencing. Some abundant sRNAs are already detectable by various staining protocols after separation on polyacrylamide gels, *e. g.*, SRP RNA (Ando *et al.*, 2002; Suzuma *et al.*, 2002) as well as the two 6S RNA homologues of *Bacillus subtilis* could be visualized after treatment with ethidium bromide.

In vitro labelling of extracted total RNA at 5' or 3' termini provides an alternative route to metabolic or *in vivo* labelling. This approach employs either T4 polynucleotide kinase and γ-[^{32}P] ATP for labelling 5' termini, or T4 RNA ligase and [^{32}P] pCp for labelling the 3' end. Labelling efficiency can vary significantly for the two termini in a given RNA pool due to secondary RNA structures which can affect the accessibility of the 5' or 3' end to be labelled, or due to the functional group at the 5' end of sRNAs. 5' RACE (rapid amplification of cDNA ends) experiments have suggested that many primary sRNA transcripts retain a 5' triphosphate which, unless removed, will preclude labelling (Argaman *et al.*, 2001; Vogel *et al.*, 2003).

Table 2.1: Systematic sRNA screens in diverse bacteria.

Bacterium	Strategy	Reference
Gram-negative bacteria		
Escherichia coli	biocomputational	Argaman *et al.*, 2001
Escherichia coli	biocomputational, oligonucleotide arrays	Wassarman *et al.*, 2001
Escherichia coli	biocomputational	Rivas *et al.*, 2001
Escherichia coli	biocomputational	Chen *et al.*, 2002
Escherichia coli	RNomics	Vogel *et al.*, 2003
Escherichia coli	Hfq co-immunoprecipitation, microarrays	Zhang *et al.*, 2003
Escherichia coli	direct cloning	Kawano *et al.*, 2005a
Salmonella typhimurium	biocomputational	Pfeiffer *et al.*, 2007
Salmonella typhimurium	biocomputational	Padalon-Brauch *et al.*, 2008
Salmonella typhimurium	Hfq co-immunoprecipitation, deep sequencing	Sittka *et al.*, 2008
Prochlorococcus marinus, *Synechococcus sp.*	biocomputational	Axmann *et al.*, 2005
Prochlorococcus marinus	microarrays	Steglich *et al.*, 2008
Borrelia burgdorferi	biocomputational	Östberg *et al.*, 2004
Vibrio cholerae	biocomputational	Lenz *et al.*, 2004
Aquifex aeolicus	RNomics	Willkomm *et al.*, 2005
Pseudomonas aeruginosa	biocomputational	Livny *et al.*, 2006
Pseudomonas aeruginosa	biocomputational, RNomics	Sonnleitner *et al.*, 2008
Sinorhizobium meliloti	biocomputational, microarrays	Ulvé *et al.*, 2007
Sinorhizobium meliloti	biocomputational	del Val *et al.*, 2007
Sinorhizobium meliloti	biocomputational	Valverde *et al.*, 2008
Caulobacter crescentus	tiling array	Landt *et al.*, 2008
Gram-positive bacteria		
Staphylococcus aureus	biocomputational	Pichon & Felden, 2005
Bacillus subtilis	biocomputational, microarray	Silvaggi *et al.*, 2006
Listeria monocytogenes	Hfq co-immunoprecipitation	Christiansen *et al.*, 2006
Listeria monocytogenes	biocomputational	Mandin *et al.*, 2007
Streptomyces coelicolor	biocomputational	Pánek *et al.*, 2008
Streptomyces coelicolor	biocomputational, sRNA cloning	Swiercz *et al.*, 2008
Archaea		
Archaeoglobus fulgidus	RNomics	Tang *et al.*, 2002
Sulfolobus solfataricus	RNomics	Tang *et al.*, 2005
Haloferax volcanii	biocomputational, RNomics	Soppa *et al.*, 2009
Haloferax volcanii	RNomics	Straub *et al.*, 2009

2.3.2. Genetic screens

Several sRNAs, such as MicF and DsrA, were identified during genetic analyses of protein factors that modulated certain physiological activities. Specifically, in a study of the genetic basis for regulation of the two *E. coli* outer membrane proteins, OmpC and OmpF (Mizuno *et al.*, 1984), it was observed that multiple-copy plasmids carrying a 300 bp DNA segment of the *ompC* promoter blocked OmpF expression. Analysis of the insert sequences of these plasmids revealed the above mentioned first member of the class of *trans*-encoded antisense RNAs, MicF RNA.

Similarly, a mucoid phenotype led to the discovery of *E. coli* DsrA RNA. When studying factors involved in capsule synthesis, such as the positive regulator RcsA, it was found that multi-copy plasmids carrying a region downstream of the *rcsA* gene caused capsule overproduction. Subcloning of this region resulted in the isolation of the noncoding *dsrA* gene. DsrA was further shown to antagonize *hns* mRNA translation by an antisense mechanism which finally explained the mucoid phenotype of multi-copy *dsrA* plasmids; under normal conditions, the histone-like protein, H-NS, silences the *rcsA* gene. Overproduction of DsrA decreases H-NS levels, which abrogates *rcsA* repression and leads to elevated capsule polysaccharide synthesis (Lease *et al.*, 1998; Sledjeski & Gottesman, 1995 and references therein). The observation that DsrA can also act as an activator of *rpoS* expression (see Section 2.1.2) initiated a genetic screen for *rpoS*-regulating sRNAs. Majdalani *et al.* (2001) transformed a pBR322-based plasmid library of *E. coli* genomic DNA fragments ranging in size from 1.5 to 5 kb (Ulbrandt *et al.*, 1997) into a strain that harboured an *rpoS::lacZ* reporter gene (translational fusion) and a mutated *dsrA* locus. Screening for colonies exhibiting enhanced β-galactosidase activity then led to the discovery of RprA RNA (Majdalani *et al.*, 2001). Similarly, the *E. coli* CsrC and the *P. fluorescens* CsrB homologue, PrrB RNA, were discovered in functional screens for genes affecting glucan biosynthesis (Aarons *et al.*, 2000; Weilbacher *et al.*, 2003). In summary, multi-copy plasmid libraries are a valuable tool for identifying sRNAs and often provide an immediate link to a physiological function. However, some sRNAs might be toxic when cloned on a multi-copy plasmid or regulate only under specific conditions and, hence, will be missed in a screen.

Besides overexpression of sRNAs based on plasmid libraries, chromosomal inactivation by random transposon insertion mutagenesis can help to find sRNA genes responsible for a certain phenotype. For example, transposon insertion mutagenesis in *Bradyrhizobium japonicum* led to the identification of the *sra* gene which encodes a 213 nt sRNA that is essential for symbiotic nodule development (Ebeling *et al.*, 1991). In addition, GlmZ, which was previously known as RyiA or SraJ RNA of hitherto unknown function (Argaman *et al.*, 2001; Wassarman *et al.*, 2001), was identified as an activator for *glmS* mRNA in a transposon mutagenesis screen for insertions that abolish the high *glmS* expression in a *yhbJ* mutant (Kalamorz *et al.*, 2007; see section 2.1.2). Fortunately, only a few sRNAs seem to be essential, namely, *E. coli* RNase P RNA and SRP RNA (see Wassarman *et al.*, 1999) as well as *Neisseria gonorrhoeae* tmRNA (Huang *et al.*, 2000). However, some sRNAs can be essential under certain stress conditions such as the DNA damage-induced *istR* locus of *E. coli*, encoding two regulatory sRNAs, cannot be deleted in SOS constitutive strains (Vogel *et al.*, 2004). Another caveat regarding gene-disruption based sRNA screening is that, due to the comparatively small size of sRNA genes, a transposon is five to ten fold more likely to disrupt a protein-coding region than an sRNA gene, assuming an average size of bacterial ORFs of 1,000 bp ($\approx$ 960 nt in *E. coli*, Blattner *et al.*, 1997).

2.3.3. Biocomputational screens

Computer-based annotation of bacterial genome sequences has been standardized by several programs for the prediction of protein-coding genes, *e. g.* `GLIMMER` (Delcher *et al.*, 1999; Salzberg *et al.*, 1998). Such programs for prediction of mRNA genes search for the longest possible reading frame and are frequently supported by the existence of orthologues in related bacteria and by

the occurrence of putative ribosome binding sites in the vicinity of the predicted start codon. As mentioned above, sRNA genes lack such characteristic identifiers and the use of computational methods to discover novel bacterial sRNAs is difficult. The ever-increasing number of completed bacterial genome sequences has greatly facilitated computer-based sRNA searches at the genomic level. The three pioneering studies that aimed at sRNA identification in *E. coli* were primarily based on comparative genomics of closely related enterobacteria such as *S. typhimurium* and *Yersinia pestis* and on prediction of orphan transcription signals in intergenic regions (Argaman *et al.*, 2001; Wassarman *et al.*, 2001; Rivas *et al.*, 2001).

The prediction strategy used by Argaman *et al.* (2001) was based on the characteristics of the ten *E. coli* sRNA genes known at the time (Wassarman *et al.*, 1999). Upon extraction of intergenic regions (IGRs) from the annotated *E. coli* genome, promoters that would match the consensus sequence recognized by the vegetative sigma factor, σ^{70}, were predicted. Furthermore, the IGRs were inspected for strong Rho-independent transcriptional terminators. *E. coli* IGRs that contained a promoter and a terminator on the same strand and within a distance of 50 - 400 bp were selected and compared to the genomes of the *Salmonella typhimurium, Yersinia pestis*, and *Klebsiella pneumoniae* by BLASTN searches. Conserved IGRs were extracted based on statistically significant alignment scores. Finally, the genomic locations of sRNA candidates were evaluated; *i. e.* putative sRNA genes that were oriented in opposite direction to both adjacent genes scored higher because these could not be conserved mRNA leaders or trailers. Overall, the prediction strategy led to 23 putative sRNA genes of which 14 could be verified as new sRNAs on Northern blots.

A similar approach was taken by Wassarman *et al.* (2001) who extracted all *E. coli* IGR sequences longer than 180 bp, compared them to *Salmonella* and *Klebsiella* IGRs, and evaluated transcription signals and sRNA gene orientation for those with a high degree of sequence conservation. Furthermore, these predictions were supported by expression analysis of putative sRNA transcripts on *E. coli* microarrays. Finally, 23 of 59 final sRNA candidates were confirmed on Northern blots. Of these, 17 were considered to be new sRNA genes. The remaining six were reclassified as small protein-coding genes based on reading frame conservation and the presence of putative Shine-Dalgarno sequences.

Rivas *et al.* (2001) introduced a conceptual change by scoring conservation of RNA secondary structure rather than of primary sequence. For this purpose, they combined structure prediction with comparative analysis of *E. coli, Salmonella* and *Klebsiella* genomes. The implemented program, QRNA, searches for mutational patterns in pairwise sequence alignments that would distinguish conserved RNA secondary structure from the background of other conserved sequence elements such as transcription factor binding sites. In contrast to the patterns of synonymous codon substitutions in conserved protein-coding regions, structural RNAs are revealed through compensatory mutations that are consistent with maintaining predicted secondary structure elements. QRNA predictions in *E. coli* yielded a total of 275 candidate RNAs. Of these final candidate genes, 49 were assayed experimentally, and 11 of these were found to express small transcripts under the single growth condition examined. Many sRNAs are known to be expressed in stationary phase or under specific stress conditions; thus, a broader set of growth conditions (Argaman *et al.*, 2001; Wassarman *et al.*, 2001) may be likely to increase the number of confirmed candidates.

The three initial screens mainly relied on conservation of sRNA genes among closely related bacteria. However, sometimes genome sequences are not yet available for related bacteria or it is desired to identify species-specific sRNAs. Another sRNA search in *E. coli* that relied solely on transcription signal prediction went into this direction (Chen *et al.*, 2002). Here, the search strategy was based on the prediction of σ^{70}-type promoter/ Rho-independent terminator pairs in IGRs that lie on the same strand within a distance of 45 - 350 bp. After removal of putative short protein-coding genes and orphan tRNA genes, short leaders, tRNA/rRNA operon fragments, or known sRNA genes, this study ended up with 144 final candidates, and seven out of eight candidates examined on Northern blots were found to be new sRNA species. Interestingly, a mere ten of the forty sRNAs known at the time were recognized by the search algorithm. An automated sRNA screening procedure for the extraction, selection and visualization of candidate IGRs has been implemented in the software package 'Intergenic Sequence Inspector', or ISI (Pichon & Felden, 2003). This program filters IGRs according to variable input parameters, including length or GC content, and can select those with significant sequence conservation among phylogenetically related bacteria. Besides all previously characterized *E. coli* sRNAs, ISI identified additional candidates in *E. coli* (Antal *et al.*, 2005) and also in the Gram-positive bacterium *Staphylococcus aureus* (Pichon & Felden, 2005).

Two additional bioinformatic approaches to identify sRNAs in *E. coli* were conducted. Carter *et al.* (2001) developed a machine learning approach that made use of neural networks and support vector machines to extract the shared features of known sRNAs for the prediction of new candidates in several prokaryotic and archaeal genomes. Similar to QRNA (Rivas & Eddy, 2001), this approach seems to be less dependent on prior knowledge of the specific RNA gene features of a given organism. The underlying algorithm employs both compositional parameters (nucleotide and dinucleotide composition) and structural motif parameters to discriminate functional RNAs from random noncoding sequences. The screen predicted ≈ 370 novel sRNA candidates in the *E. coli* genome which await experimental verification. More recently, boosted genetic programming was used to create sRNA classifiers in order to select noncoding functional RNA sequences from intergenic sequences which predicted several new sRNA candidates (Saetrom *et al.*, 2005).

Biocomputational approaches were also successfully applied to scan the genomes of entirely unrelated bacteria, including four marine cyanobacteria of the *Prochlorococcus-Synechococcus* lineage (Axmann *et al.*, 2005). It has been shown that thermodynamic stability values derived from the consensus folding of aligned related sequences allow effective prediction of functional RNAs (Washietl & Hofacker, 2004; Washietl *et al.*, 2005). Based on this strategy, Axmann *et al.* (2005) scored alignments of IGRs that are conserved among three *Prochlorococcus* genomes and one *Synechococcus* genome using `ALIFOLDZ` (Washietl & Hofacker, 2004; Washietl *et al.*, 2005). Expression analysis of the highest scoring candidate regions under various growth and stress conditions confirmed seven new sRNAs in *Prochlorococcus marinus*, several of which had homologues in the other three strains. In addition, these searches also uncovered new cyanobacterial 6S RNA orthologues, *i. e.* in addition to the 6S RNA-like genes previously reported in other *Synechococcus* strains (Watanabe *et al.*, 1997).

The successful use of bioinformatics to identify sRNAs in bacteria other than *E. coli* has been limited by the lack of well-defined promoter consensus sequences in most species. However, several groups have recently demonstrated that sRNA genes can be predicted even without putative pro-

moters or transcription factor binding sites. For example, the search strategy of Waldor group used only sequence conservation of IGRs and predictions of Rho-independent terminators. This led to the successful identification of several sRNAs in *Vibrio cholerae* and *Pseudomonas aeruginosa* (Livny *et al.*, 2005; Livny *et al.*, 2006). Furthermore, the same group conducted genome-wide annotations for putative sRNA-encoding genes in the intergenic regions of eleven diverse pathogens using their `sRNAPredict2` program (Livny *et al.*, 2006). In total, they predicted more than 2700 previously unannotated candidate sRNA loci in diverse bacteria which, however, await experimental verfication. Recently, the extension to kingdome-wide predictions and annotations of intergenic sRNA-encoding genes in the 932 bacterial replicons in the NCBI database predicted more than 45,000 candidate loci (Livny *et al.*, 2008). Though, how many of the candiates are real sRNAs remains to be tested.

Despite this high number of predicted candidates, probably diverse sRNA genes have been missed in predictions. For example, each of the existing sRNA gene finders is unable to identify all the experimentally validated sRNA genes, indicating that all these *in silico* methods have limitations. Furthermore, most of the prediction strategies were initially designed for predictions in *E. coli* with serious limitations and the need for adjustments to apply them to other genomes. Moreover, they are often limited to intergenic regions. Recent studies using a hidden Markov model (Yachie *et al.*, 2006) enabled identification of sRNAs in protein-coding regions, but their efficiencies need to be improved.

2.3.4. Identification of sRNAs by transcription factor binding sites

Several sRNAs have been identified due to the presence of specific transcription factor binding sites in their promoter regions. For example, an sRNA search in *Vibrio* genomes used many features of the prior *E. coli* screens and added yet another layer (Lenz *et al.*, 2004). Based on some previous observations, the Bassler group assumed that one or more unknown sRNAs could be responsible for an Hfq-dependent regulation of the quorum sensing master regulator, LuxR. As further results indicated that such sRNAs would be activated by the sigma factor σ^{54}, a computer-based method was developed to scan *Vibrio cholerae* IGRs for pairs of σ^{54} binding sites and Rho-independent terminators. Combination with conservation analysis in two other *Vibrio* species led to the discovery of four novel sRNAs that are almost identical and conserved in all three investigated *Vibrio* species.

Similarly, *Pseudomonas* homologues of the *E. coli* RyhB RNA were discovered by functional evidence that pointed to the involvement of a hidden sRNA (Wilderman *et al.*, 2004). In many bacteria, iron homeostasis is controlled primarily by the ferric uptake regulator (Fur), a transcriptional repressor. However, some genes, including those involved in iron storage, are positively regulated by Fur. *E. coli* RyhB RNA was found to be repressed by Fur, and derepression of the sRNA under iron-limiting conditions allows downregulation of several mRNAs by an antisense mechanism (Massé & Gottesman, 2002). The *ryhB* gene, along with its promoter and Fur binding site, is well conserved in enterobacteria but could not be found in *Pseudomonas aeruginosa*, another organism in which positive regulation by Fur had been observed and left unexplained (Ochsner *et al.*, 2002). However, a biocomputational approach based on pattern searches for Fur-consensus binding site in IGRs combined with predictions for Rho-independent terminators identified two RyhB homologues in *P. aeruginosa* (Wilderman *et al.*, 2004). These sRNAs are more than 95% identical to each other

and are expressed in a Fur-dependent manner. In addition, PrrF1 and PrrF2 RNAs show a stretch of complementarity with the translation initiation region of *sodB* mRNA, further suggesting that these newly discovered sRNAs are functional homologues of RyhB.

In *Streptococcus pneumoniae* five sRNAs were fortuitously identified during analysis for CiaR binding sites, which is part of the two-component regulatory system CiaRH (Halfmann *et al.*, 2007). These small RNAs, designated csRNAs for *cia*-dependent small RNAs, are non-coding, between 87 and 151 nt in size, and show a high degree of similarity to each other. This suggests that systematic screens for transcription factor binding sites could be used to identify novel sRNAs. In addition this would immediately provide a link to the expression conditions of the sRNA and give a hint of its function.

2.3.5. RNomics

Shotgun cloning of RNA (also termed experimental RNomics) led to the discovery of hundreds of noncoding RNAs in several eukaryotes and archaebacteria (see, for example, Hüttenhofer *et al.*, 2001; Marker *et al.*, 2002; Tang *et al.*, 2002, 2005; Yuan *et al.*, 2003). Typical RNomics protocols include a size-fractionation step of total RNA on polyacrylamide gels, followed by directional cDNA cloning of the gel-extracted RNA and sequencing of the resulting libraries. This method of randomly cloning as many small RNA fragments as possible aims to comprehensively identify RNAs that are expressed from a given genome under a given set of conditions, irrespective of whether they are primary or processed transcripts.

In the two shotgun cloning studies conducted in *E. coli* to date the RNA size range was either 50 - 500 nt (Vogel *et al.*, 2003) or 30 - 65 nt (Kawano *et al.*, 2005a). Following size-fractionation, individual cDNA libraries representing two or three distinct growth phases were constructed, based on earlier observations that many *E. coli* sRNAs are expressed in a growth-rate specific manner (*e. g.* Argaman *et al.*, 2001; Wassarman *et al.*, 2001). In the first study, extracted RNA was C-tailed with poly(A) polymerase, followed by reverse transcription and construction of cDNA libraries (Vogel *et al.*, 2003). Individual library clones (10,000) were spotted on high-density filters and hybridized with a mix of rRNA and tRNA probes to exclude such clones from further study, and cDNA clones that passed this test were sequenced. This identified, in addition to sRNA genes, fragments of mRNA leaders and trailers that accumulated as distinct sRNAs, as well as many sequences derived from within coding regions. Because cDNA cloning was directional, several of the sequences could be classified as antisense transcripts of mRNA coding regions.

The cloning strategy of Kawano *et al.* (2005a) was more similar to that used to discover eukaryotic miRNAs (Elbashir *et al.*, 2001; Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Lee & Ambros, 2001). Here, the extracted small-sized RNA fragments were ligated to specific 5' and 3' RNA adapter molecules, reverse-transcribed, and PCR amplified. Prior to cloning, the amplification products were concatenated in order to increase the sequence information per individual library clone. This led to the detection of 5' UTR-derived and 3' UTR-derived sRNAs as well as *cis*-encoded antisense RNAs encoded opposite to protein coding genes. Beyond *E. coli*, a bacterial RNomics screen was conducted in *Aquifex aeolicus*, a hyperthermophilic bacterium (Willkomm *et al.*, 2005). Shotgun cloning enabled the detection of about half a dozen sRNAs candidates, some from the intergenic

space and some that were antisense RNAs. With the exception of housekeeping sRNAs, these candidates were the first to be described in hyperthermophilic eubacteria and included the *A. aeolicus* 6S RNA homologue.

A major limitation of the RNomics approach is the need to evaluate a large number of sequences and an overrepresentation of the highly abundant rRNAs and tRNAs in the cDNA libraries. Furthermore, less structured RNAs are more easily transcribed into cDNA and, hence, could be overrepresented in the libraries. Expression of most sRNAs is often limited to specific growth conditions. Thus, it will remain difficult to select all possible growth conditions to recover all sRNAs encoded in a bacterial genome. The strength of cloning-based approaches is their ability to identify sRNAs from intergenic regions that are not conserved in species related to *E. coli*, *e. g. Salmonella*, since such candidates would not rank highly in screens having sRNA gene conservation as the main criterion. Furthermore, they allow detection of *cis*-encoded antisense RNAs which were a prominent class of small RNA molecules found by Kawano *et al.* (2005b). The methodology could also be improved in some aspects, *e. g.*, by affinity-based removal of rRNA from the total RNA pool prior to cloning. In addition, the combination of RNomics with the recently developed high-throughput sequencing technologies will facilitate analysis of cDNA libaries in more depth. Furthermore, it avoids the bacterial cloning step and allows parallel sequencing of libraries from different growth conditions. The use of deep sequencing technologies to analyse cDNA libraries and recent studies which employ this approach will be discussed in Chapters 5 and 6.

2.3.6. Homology searches

Since many of the global *E. coli* sRNA screens took advantage of sequence conservation in related enterobacteria (Argaman *et al.*, 2001; Wassarman *et al.*, 2001; Rivas *et al.*, 2001), most of the sRNAs that were identfied in *E. coli* have also homologues in related species such as *Salmonella* and *Yersinia* (Hershberg *et al.*, 2003; Papenfort *et al.*, 2008). However, comprehensive BLAST analysis of more than one hundred completed bacterial genomes did not yield significant sequence similarity for most of these 55 *E. coli* sRNAs beyond *Yersinia pestis* (Griffiths-Jones *et al.*, 2005; Hershberg *et al.*, 2003). Conservation of sRNA flanking genes decreases with phylogenetic distance which, in turn, suggests that an sRNA gene is more likely to be found in distantly related bacteria when both flanking genes are also conserved. The observed correlation between sRNA gene conservation and the evolutionary distance of organisms was subsequently confirmed by an independent study (Zhang *et al.*, 2004). Interestingly, sRNA and protein-coding genes were found to evolve at the same rate in bacteria. In contrast, tRNA genes tend to be more conserved than other genomic regions, and loci of unknown function evolve much faster than the average.

Recent studies indicate that functional sRNAs often lack significant sequence similarity. For example, the *E. coli* CsrB and CsrC RNAs show only little sequence conservation to their functional homologues in *Vibrio* (Lenz *et al.*, 2005). Similarly, the functional homologues of the Fur-regulated sRNA RyhB in *E. coli* (Massé & Gottesman, 2002) and *Pseudomonas* (Wilderman *et al.*, 2004) share only little sequence similarity. Moreover, the *V. cholerae* RyhB is more than twice as long as the *E. coli* RyhB and the actual sequence similarity between *E. coli* and *Vibrio* is limited to a central 28 bp region (Davis *et al.*, 2005).

2.3.7. Microarray detection

Microarrays have become the method of choice for monitoring mRNA expression profiles at the genome level and were also successfully used for studying sRNA expression or even for finding new sRNA transcripts. However, as most of the commercially available arrays are limited to ORFs and at best include tRNA and rRNA genes, custom arrays which include probes specific to IGRs often have to be designed. Selinger *et al.* (2000) introduced a high-density oligonucleotide probe array for *E. coli* that covers also IGRs of > 40 bp in addition to strand-specific probes for all mRNA, tRNA, and rRNA regions. Although this study primarily focused on the sensitivity and reproducibility of mRNA level profiling, it provided preliminary data on some intergenic and antisense RNAs that were secondarily detected.

Similarly, Tjaden *et al.* (2002) used a whole genome array and identified more than 1,000 additional transcripts in the intergenic regions of the *E. coli* genome and further classification revealed ≈ 300 novel transcripts with unknown function. The same study, however, emphasizes the need to validate initial microarray results by independent experimental techniques such as Northern hybridization or RNase protection assays. Wassarman *et al.* (2001) supplemented their biocomputational sRNA prediction by using the same type of array and specific analysis of the intergenic transcriptional output.

The combination of microarrays with comparative genomics has led to the identification of diverse sRNAs in *E. coli* and *Staphylococcus aureus*, as well as sRNAs under sporulation control in *Bacillus subtilis* (Wassarman *et al.*, 2001; Pichon & Felden, 2005; Silvaggi *et al.*, 2006). Recently, several novel *Caulobacter crescentus* sRNAs could be identified by analysis of RNA expression levels assayed using a tiled *Caulobacter* microarray and a protocol optimized for detection of sRNAs (Landt *et al.*, 2008). In the cyanobacterium *Prochlorococcus*, analysis of RNA from different stress conditions on Affymetrix microarrays revealed in addition to several new sRNA genes and antisense transcripts many new short mRNAs that are expressed from intergenic regions (Steglich *et al.*, 2008). Moreover, the sRNAs were found to be clustered in previously identified genomic islands which carry genes of significance to the ecology of this organism.

Although microarrays have been successfully used for sRNA identification, they have some disadvantages. These include the limited availability of high-density arrays for diverse bacteria or problems inherent to labelling and cDNA synthesis of small structured RNAs to be used as samples on microarrays. These RNAs could easily be missed on arrays, especially if the resolution of probes is not high enough. Recent studies report the use of alternative RNA detection methods which could circumvent this problem (Hu *et al.*, 2006; Zhang *et al.*, 2003). In one of these approaches, Zhang *et al.* (2003) immunoprecipitated the Hfq protein (see below, Section 2.3.8) and directly detected the sRNAs bound to this protein on microarrays with high-affinity and nucleotide sequence-independent antibodies specific for RNA:DNA hybrids. The same antibody was used by Hu *et al.* (2006) to detect unmodified RNA which was hybridized directly to DNA microarrays. The antibody-based method turned out to detect low abundance small RNAs from *E. coli* much more efficiently than the commonly-used cDNA-based method.

2.3.8. Co-purification with proteins

Many sRNAs are assumed to form RNP complexes with proteins. These include RNA-binding proteins which help an sRNA to fold into its active conformation, shield it from nucleases prior to reaching a target, or promote its annealing with target mRNAs, such as Hfq (see Section 2.2.1). Other sRNAs interact with proteins to directly regulate their activity (see Section 2.1.3). Therefore, several sRNAs have been discovered by co-purification with abundant RNA-binding proteins. These include CsrB RNA of *E. coli* and RsmZ RNA of *Pseudomonas fluorescens* which were co-purified with their target proteins CsrA and RsmA, respectively (Heeb *et al.*, 2002; Liu *et al.*, 1997). As mentioned before, the genes of the CsrB-like RNA family show poor sequence similarity. Hence, co-purification with interacting proteins could be an alternative strategy for finding sRNAs whose primary sequences have diverged to an extent that they are unidentifiable by BLASTN searches.

To date, Hfq is the best characterized sRNA binder and has been suggested to interact with more than one third of the sRNAs known in *E. coli* (Valentin-Hansen *et al.*, 2004; Zhang *et al.*, 2003). Immunoprecipitation of Hfq followed by microarray detection of the RNAs bound to the protein was recently employed as a new screening approach. In this way, half a dozen new sRNAs were found in *E. coli* (Zhang *et al.*, 2003). Similar approaches for finding Hfq-binding sRNAs have been taken in the bacterial pathogens *Listeria monocytogenes* (Christiansen *et al.*, 2006) and *Pseudomonas aeruginosa* (Sonnleitner *et al.*, 2008). The latter study constructed cDNA libraries of RNAs which were co-immunoprecipitated by Hfq and subjected them to shotgun-cloning, whereas the Hfq-binding RNAs were identified by enzymatic RNA sequencing. Since species-specific Hfq antibodies may not always be available, affinity tags like FLAG or $(His)_6$ provide epitopes that can be targeted by specific antibodies.

A genomic SELEX approach which would in principle cover all sRNAs that are encoded by a given genome was recently reported by the Schroeder lab (Lorenz *et al.*, 2006). Here, a representative library of the *E. coli* genome was constructed from random sequences of 50 - 500 bp (Singer *et al.*, 1997). These fragments were *in vitro* transcribed with T7 RNA polymerase, incubated with Hfq, and selected for Hfq binding on filters. This approach has the advantage that it circumvents the need of an sRNA to be expressed under the examined conditions as it generates RNAs *in vitro*.

2.4. Target identification and verification of bacterial sRNAs

The previous section introduced diverse approaches for the identification of sRNAs and indicated the presence of many sRNA genes in bacterial genomes which await functional characterization. For this purpose, diverse experimental and biocomputational methods have been developed to identify the cellular targets of sRNAs (for reviews see, *e. g.*, Pichon & Felden, 2008; Vogel & Wagner, 2007). However, identification of sRNA targets is only the first step in functional elucidation of an sRNA, and *in vivo* validation of target regulation is a crucial step. This can be achieved by several approaches.

Generally, sRNA-mediated alteration of mRNA or protein levels can be monitored on Northern or Western blots, respectively. Sometimes antibodies are not available for a designated target and, therefore, chromosomal or plasmid borne translational reporter gene fusions, for example to *lacZ*,

which encodes β-galactosidase, have been used to examine target regulation. Recently, a reporter-gene system based on green fluorescent protein (GFP) was developed to study sRNA-target interactions in *E. coli* and also distant bacteria (Urban & Vogel, 2007). This system is based on two plasmids in which both the sRNA and target-mRNA fusion are expressed under control of a constitutive promoter and are, thus, uncoupled from any transcriptional control. This system provides a convenient way to examine post-transcriptional target regulation and gives hints whether targets are regulated directly. However, the most appreciated proof for a direct sRNA-mRNA interaction are compensatory base-pair exchanges. Mutations in either the interaction site of the sRNA or the target site should abolish regulation, whereas compensatory mutations in both sites should restore regulation (Altuvia *et al.*, 1998; Bouvier *et al.*, 2008; Kawamoto *et al.*, 2006; Majdalani *et al.*, 1998; Papenfort *et al.*, 2008; Udekwu *et al.*, 2005; Vogel *et al.*, 2004).

Furthermore, *in vitro* experiments can help to validate interactions between two RNAs and provide insights into the mechanism of action of sRNA-mediated regulation. These biochemical experiments include determination of binding affinities by gel-shift assays of *in vitro* transcribed RNAs or exact mapping of interaction sites by footprinting experiments using enzymatic probing or lead(II)-induced cleavage (reviewed in Chevalier *et al.*, 2009). Moreover, sRNA-mediated inhibition of translation initiation can be investigated in ribosome toeprinting assays which monitor formation of the ternary initiation complex on an *in vitro* transcribed mRNA fragment (Altuvia *et al.*, 1998; Chen *et al.*, 2004; Hartz *et al.*, 1988; Huntzinger *et al.*, 2005; Udekwu *et al.*, 2005). Overall, there are a variety of possible strategies to validate putative mRNA targets. These can, in addition, help to assess the specificity of the diverse approaches for target identification introduced below.

2.4.1. Experimental approaches

2.4.1.1. Proteomics

Perhaps the most direct way to identify putative targets of bacterial sRNAs is to analyse protein patterns in an sRNA deletion or overexpression strain on 1D- or 2D-gels. Bands or spots of differentially expressed proteins can be excised from the gels and analysed by mass spectrometry. For example, comparison of an *E. coli* wild-type strain to an isogenic *gcvB* mutant strain revealed changes in the levels of several ABC transporter proteins (Urbanowski *et al.*, 2000). Specifically, the *gcvB* mutant shows increased levels of the periplasmic transporter proteins OppA, DppA, GltI and LivK. Similarly, OmpA turned out to be upregulated in deletion mutant of the sRNA VrrA in *V. cholerae* when whole-cell protein profiles were compared by SDS-PAGE (Song *et al.*, 2008). For some sRNAs, deletion is not enough to reveal changed proteins and, in fact, overexpression is necessary to induce changes in the protein patterns. In *E. coli*, this strategy identified for example OmpA as a target of MicA RNA upon overexpression of the sRNA from a multicopy plasmid (Rasmussen *et al.*, 2005; Udekwu *et al.*, 2005). Similarly, overexpression of Spot42 RNA reduced synthesis of the GalK protein, which is translated from the polycistronic *galETKM* mRNA (Møller *et al.*, 2002), or the GlmS protein is upregulated when GlmY RNA is overexpressed (Urban *et al.*, 2007).

One limitation of the proteomics approach is that based on either reduced or increased protein level when the sRNA is lacking or is overexpressed it can not be distinguished whether the regulation

is direct, *i. e.* interaction between the sRNA and the mRNA encoding the protein, or indirect via additional regulators. Furthermore, the number of individual protein spots or bands that can be resolved is relatively low. However, if an sRNA inhibits only translation and does not lead to mRNA degradation, regulated genes can still be identified by proteomics due to altered protein levels but not based on analysis of mRNA levels on microarrays.

2.4.1.2. Microarrays

The sRNA-mediated activation or repression of translation is often accompanied by changes in mRNA levels. This can be explained by changes of mRNA stability due to recruitment of RNases such as RNase E or RNase III upon duplex formation or simply higher accessibility for RNases on mRNAs that are less associated with actively translating ribosomes. Therefore, several approaches have used microarrays to monitor mRNA changes upon sRNA regulation. For example, comparison of mRNA extracted from strains carrying a control vector and an overexpression vector for DsrA RNA on filter-based DNA arrays revealed several genes that are up- or downregulated (Lease *et al.*, 2004). Similar to analysis of global changes in protein patterns, this method is not able to distinguish between directly and indirectly regulated targets. For this purpose, 'pulse'-expression of a given sRNA from an inducible promoter, such as the tightly controlled arabinose-inducible pBAD promoter (Guzman *et al.*, 1995), was used recently. This strategy is likely to avoid the pleiotropic effects that can be expected to result from constitutive sRNA expression. A pulse of induction and analysis of mRNA changes on microarrays after a short time interval (typically 10 to 15 min) is likely to reveal mainly direct targets and avoid downstream effects. These include alterations of additional targets after longer induction times if, *e. g.*, a transcriptional regulator is a direct target of the sRNA. Several sRNAs from *E. coli* and *Salmonella* have now been characterized using this approach. In one of these initial studies, genes regulated by RyhB RNA in *E. coli* were identified by pulse-expression of RyhB from a strong inducible promoter and scoring of changes in global mRNA levels on whole-genome microarrays. For RyhB, this method predicted 18 additional target mRNAs encoding for 56 proteins most of them with documented functions in iron metabolism (Massé *et al.*, 2005) making RyhB a paradigm for a conserved principle of sRNA-mediated control in a given physiological circuit, here iron homeostasis. The same approach was successfully used to identify RybB RNA as a factor that selectively accelerates the decay of multiple major outer membrane protein (OMP) mRNAs upon induction of the envelope stress response in *Salmonella* (Papenfort *et al.*, 2006). Furthermore, two tandem oriented, nearly identical sRNAs, OmrA and OmrB, were found to repress an overlapping set of mRNAs that encode outer membrane proteins (Guillier *et al.*, 2006).

2.4.1.3. Co-immunoprecipitation of direct interaction partners

Many of the bacterial sRNAs require the activity of the RNA chaperone Hfq which facilitates the often imperfect base-pairing interactions between sRNA and mRNA (see Section 2.2.1). Therefore, co-immunoprecipiation of RNA bound to Hfq could reveal potential mRNA targets. Co-immunoprecipitation with Hfq-specific antisera and direct detection of the bound RNAs on genomic high-density oligonucleotide microarrays were combined in an initial global study in *E. coli* (Zhang *et al.*, 2003). Although this method successfully detected diverse sRNAs and mRNAs in *E. coli*, similar studies in other bacteria are limited by the requirement of high-density microarrays and spe-

cialized antibodies. An alternative approach identified individual abundant Hfq-associated RNAs by cDNA cloning or direct sequencing (Antal *et al.*, 2005; Christiansen *et al.*, 2006); however, these methods are not appropriate for large-scale analyses. Although the co-immunoprecipitation strategy recovers putative sRNA-regulated mRNAs, the actual sRNA-mRNA pairs remain unclear.

Two groups have tried to capture direct mRNA interaction partners using the sRNA as a bait. The first study immobilized complexes of His-tagged Hfq and RydC RNA on a nickel column and incubated them with the pool of extracted cellular mRNA (Antal *et al.*, 2005). RNAs that were retained on the column, were eluted, converted into cDNA and sequenced. This led to the identification of the polycistronic *yejABEF* mRNA which encodes a putative ABC transporter as a target of RydC. The second study used *in vitro* transcribed and biotinylated RseX RNA, bound to streptavidin magnetic beads, to capture mRNAs from *E. coli* total RNA extracts (Douchin *et al.*, 2006). Affinity captured RNAs were converted to cDNA, fluorescently labelled, and hybridized onto *E. coli* whole-genome microarrays. This revealed the outer membrane protein encoding mRNAs *ompA* and *ompC* as RseX targets.

Capture of direct interaction partners can also be applied to RNAs which sequester protein activity. For example, RNA affinity chromatography was used to identify tryptophanase as the direct target of the plasmid-encoded Rcd RNA (Chant & Summers, 2007; see section 2.1.3). *In vitro* transcribed Rcd RNA was cross-linked to sepharose, incubated with *E. coli* cell lysates, and MALDI spectroscopy of retained proteins revealed TnaA as a direct interaction partner of Rcd. A recent study purified and identified sRNA-binding factors from *E. coli* via affinity chromatography of aptamer tagged sRNAs and mass spectrometry (Windbichler *et al.*, 2008). This recovered RNA polymerase beta-subunit, host factor Hfq and ribosomal protein S1 as sRNA-binding proteins in addition to several other factors. Also in *Salmonella*, RNA-based affinity chromatography can be successfully used to purify sRNAs following their expression as aptamer-tagged variants *in vivo* and to co-purify Hfq (Said *et al.*, 2009, submitted). These affinity purification strategies should facilitate the isolation of *in vivo* assembled sRNA-protein complexes in a wide range of bacteria.

2.4.2. Biocomputational target predictions

The enterobacterial sRNAs exhibit short and/or imperfect complementarity to their target(s). Figure 2.4 gives two examples for experimentally verfied sRNA-target interactions. Nine residues of RyhB sRNA interact with a region overlapping the start codon of *sodB* mRNA (Geissmann & Touati, 2004), and OxyS sRNA targets *fhlA* mRNA through the formation of two short kissing complexes of nine and seven basepairs, respectively (Argaman & Altuvia, 2000). Systematic mutational analysis of SgrS-*ptsG* RNA duplexes revealed that six residues of this interaction are key to mediating SgrS repression of *ptsG* mRNA (Kawamoto *et al.*, 2006). This limited sequence complementarity has rendered the identification of new sRNA targets difficult, and for the majority of the sRNAs studied to date, a single mRNA remains the only experimentally validated target.

For some sRNAs, targets could be identified by simple `BLASTN` searches. These include interactions between MicC and *ompC* mRNA as well as IstR-1 with *tisAB* mRNAs which form rather long duplexes (Chen *et al.*, 2004; Vogel *et al.*, 2004). However, to identify short and imperfect interactions more complicated algorithms are required. These could include features such as minimization of hybridization energies, minimal lengths for seed interactions, conservation of the interaction in

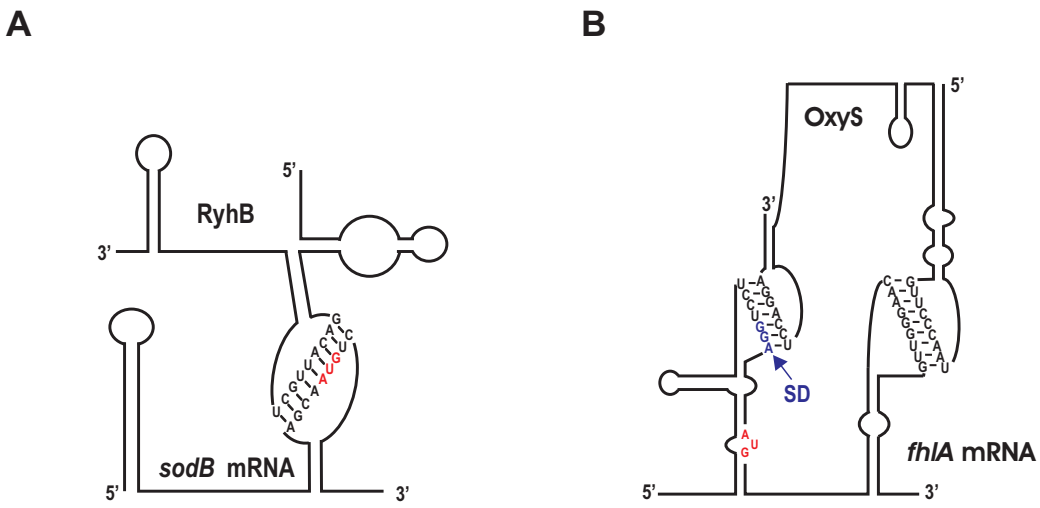


Figure 2.4: Examples for sRNA-target mRNA interactions. (A) RyhB-*sodB* mRNA. RyhB interacts with its target *sodB* mRNA by base pairing across the start codon and thereby inhibits translation (Geissmann & Touati, 2004). The interaction between the two RNAs is mediated by Hfq, which induces a structural rearrangement of the *sodB* 5'-UTR. **(B)** OxyS-*fhlA* mRNA. OxyS represses translation of *fhlA* mRNA by the formation of a loop-loop kissing complex (Argaman & Altuvia, 2000). The complex involves two interaction sites of 7 and 9 bp. One site in OxyS mediates base-pairing with a short sequence overlapping the ribosome binding site and a second site pairs with residues further downstream within the coding region of *fhlA* mRNA. Start codons are indicated in red, the SD sequence of *fhlA* mRNA in blue, respectively.

related bacteria or location of the target site around the RBS and start codon. While several approaches have been described for identifying targets of miRNA genes in eukaryotes (Brennecke *et al.*, 2005; Grimson *et al.*, 2007; Krek *et al.*, 2005; Rajewsky, 2006; Rehmsmeier *et al.*, 2004), only a few programs have been developed for the prediction of bacterial sRNA regulators. One approach that was successfully applied to the *micA-ompA* regulation searches for sRNA complementarities in sequence windows containing translation initiation sites, allowing noncontiguous pairing, and the candidate target is compared to reiterated searches in related bacteria (Udekwu *et al.*, 2005). Another computational tool was developed for the prediction of sRNA targets in *Listeria monocytogenes* and a set of validated sRNA-target mRNA pairs was used as training set (Mandin *et al.*, 2007). Furthermore, the program scores thermodynamic pairing energies between an sRNA and putative mRNA targets, and genes are selected according to their abilities to pair with the sRNA around the translational start and stop sites. This algorithm predicted several targets in *L. monocytogenes*; however, it is not available as a webserver.

The only web-based prediction tool that is so far available for bacterial target predictions is `TargetRNA` (Tjaden, 2008). This program calculates optimal hybridization scores between an sRNA and all the mRNAs from a given genome, focussing around the translational start sites (Tjaden *et al.*, 2006). The underlying model for hybridization scoring is based on the `RNAhybrid` algorithm which was developed for the prediction of miRNA targets (Rehmsmeier *et al.*, 2004). Furthermore, `TargetRNA` allows to investigate orthologous sRNA-mRNA interactions. The performance of `TargetRNA` has been validated experimentally in *E. coli* using Northern Blots and

microarray analysis (Tjaden *et al.*, 2006). Several other programs have been developed to predict RNA-RNA interactions which could be helpful to identify sRNA targets. These include algorithms for the prediction of secondary structure of two interacting RNA molecules by free energy minimization (Alkan *et al.*, 2006), or the `RNAup` software which allows to determine binding free energies of short oligomers to mRNA targest by computing probabilities of a sequence interval to be unpaired (Mückstein *et al.*, 2006). `RNAcofold` computes the hybridization energy and the base-pairing pattern of a pair of interacting RNA molecules and, in contrast to earlier approaches, accounts for complex internal structures in both RNAs (Bernhart *et al.*, 2006). Furthermore, it was suggested that other criteria such as target site accessibility and seed interactions could improve predictions (Busch *et al.*, 2008; Tafer & Hofacker, 2008). Zhang *et al.* (2006) described a new algorithm for identifying negatively regulated targets, which incorporates five different features: structure of the sRNAs, Hfq-binding sites, the 5' end of mRNAs, loop-centred extension alignments, and conservation profiles.

Despite this number of different approaches, the likelihood of detecting all the mRNA targets of a given sRNA by computational methods is still low as several known interactions could not be predicted by these tools. Furthermore, these computational methods still have only low specificity. In this regard, a combined approach of experimental identification and biocomputational prediction of interactions would probably be most effective in finding sRNA targets. Several proteins have been shown to be closely associated with bacterial sRNA function and structures (reviewed in Pichon & Felden, 2007). In contrast to the majority of enterobacterial sRNAs which act as antisense regulators, some sRNAs have been found to directly interact with proteins and antagonize their activity (see Section 2.1.3). Currently, these interactions can be detected only by experimental methods (Section 2.4.1.3). Therefore, the computational prediction of direct sRNA-protein interactions is one of the next challenges.

2.5. Multiple targeting by sRNAs in bacteria

Although the number of functionally characterized sRNAs and biochemically verified sRNA-target interactions is still lagging behind the number of newly identified sRNAs, there is growing evidence that sRNAs regulate rather multiple targets than individual mRNAs and thereby reprogram gene expression at the post-transcriptional level (reviewed in Papenfort & Vogel, 2009, *in press*). For example, OxyS sRNA was early observed to alter the steady-state levels of $>$40 abundant *E. coli* proteins (Altuvia *et al.*, 1997). Similarly, pioneering work on *E. coli* DsrA revealed that such multiple target regulation can be direct, *i. e.* this sRNA directly interacted with more than one mRNA (Lease *et al.*, 1998; Majdalani *et al.*, 1998). Nevertheless, evidence for direct interactions of sRNAs with multiple mRNAs has been rare in enterobacteria and has come primarily from investigations of RNAIII of the Gram-positive pathogen *Staphylococcus aureus* (Boisset *et al.*, 2007). As previously mentioned, RNAIII is a bifuncional molecule that encodes δ-hemolysin in its 5' part while it also acts as a noncoding regulatory RNA (Section 2.1.6). Many of the mRNAs demonstrated to be direct targets of *S. aureus* RNAIII encode bacterial virulence factors and a short ($\approx$ 40-nt) region within the 514-nt RNAIII facilitates base-pairing to multiple mRNAs (Boisset *et al.*, 2007; Geisinger *et al.*, 2006; Huntzinger *et al.*, 2005).

Recently, several new targets with functions in iron metabolism of the Fur-repressed RNA RyhB were identified on microarrays upon pulse-expression of the sRNA (see Section 2.4.1.2). How many of these genes are directly regulated by RyhB is yet to be seen. At least for three targets, *sodB*, *uof-fur*, and *shiA*, base-pairing interactions have been experimentally verified so far (Geissmann & Touati, 2004; Prévost *et al.*, 2007; Vecerek *et al.*, 2007). Some of the targets are repressed by RyhB (*sodB*, *uof-fur*), whereas *shiA* is activated.

In bacteria, the σ^E (envelope stress sigma factor) response counteracts the accumulation of unfolded outer membrane proteins in the periplasm, and two sRNAs, MicA and RybB, were found to halt OMP synthesis when porin production threatens outer membrane homeostasis within this regulon (Papenfort *et al.*, 2006). In this circuit, the 80 nt RybB sRNA, which is transcriptionally induced by σ^E (Papenfort *et al.*, 2006), is the most globally acting OMP regulator of *Salmonella* as revealed upon pulse-expression and analysis of mRNA changes on microarrays (see Section 2.4.1.2 and Papenfort *et al.*, 2006). Repression of several OMPs is mediated by base-pairing with the 5' UTRs or coding regions of *omp* target mRNAs; these RNA interactions are generally short and imperfect, and involve the conserved 5' end of RybB sRNA (Bouvier *et al.*, 2008; Mika *et al.*, 2009, submitted). Pulse-expression combined with microarray analysis revealed that two additional, nearly identical sRNAs, OmrA and OmrB, repress an overlapping set of mRNAs that encode OMPs, *e. g.*, the outer membrane protease, OmpT, and the specific gated channels for iron-siderophore complexes, CirA, FecA, and FepA. (Guillier *et al.*, 2006). Specifically, the first nine nucleotides from the 5' end, which are conserved beyond *E. coli* and *Salmonella*, were recently shown to harbour critical residues to directly recognize multiple target mRNAs (Guillier & Gottesman, 2008). All these recent studies indicate that diverse sRNAs are able to directly regulate multiple target mRNAs. However, the understanding of the underlying mechanisms is just beginning. One of these multiple target regulators, GcvB RNA, which is conserved in *E. coli* and *Salmonella*, will be investigated in Chapters 3 and 4 of this thesis.

2.6. Model pathogens used in this study

2.6.1. *Salmonella enterica* Serovar Typhimurium

Salmonella is a Gram-negative bacterium closely related to *E. coli* K12. Unlike the latter, which are non-pathogenic, *Salmonella* species are ubiquitous human and animal pathogens that cause a variety of food-borne infections such as gastroenteritis or typhoid fever. To invade and replicate in eukaryotic cells, *Salmonella* relies upon a range of laterally acquired virulence regions, the so-called *Salmonella* pathogenicity islands (SPIs). Of these, SPI-1 and SPI-2 encode type 3 secretion systems (T3SS), which translocate effector proteins to facilitate either invasion of non-phagocytic cells (SPI-1) or survival within macrophages (SPI-2). The secreted effectors are encoded by SPI-1 or SPI-2, by other minor SPIs, or by individual genes scattered throughout the *Salmonella* chromosome. Sequence analysis of *Salmonella typhi* and *Salmonella typhimurium* genomes revealed the presence of many insertions compared to the *E. coli* genome. More than 25 % of the total genetic material has been laterally acquired since *Salmonella* diverged from *E. coli* (Porwollik & McClelland, 2003). The evolutionarily close relationship with *E. coli* and the pathogen-specific aspects make *Salmonella* an attractive candidate for RNA research. The auxiliary proteins (nucleases,

RNA chaperones) typically involved in RNA-based circuits are highly conserved between the two species. Besides conserved sRNAs from *E. coli*, the *Salmonella*-specific regions could encode for new sRNAs or other RNA elements whose function would be missed in *E. coli*.

2.6.1.1. Hfq phenotype

Perhaps the strongest evidence that sRNAs could have important functions in *Salmonella* is derived from work on the RNA chaperone Hfq. The numbers of phenotypes and deregulated genes observed in an *hfq* deletion mutant of *Salmonella* exceed those reported for any other pathogen. Deletion of *hfq* has for long been known to impair the expression of σ^S (Brown & Elliott, 1996), a general stress sigma factor essential for *Salmonella* virulence in mice (Fang *et al.*, 1992). More recently, an *hfq* mutation was shown to attenuate invasion of *Salmonella* into epithelial cells, secretion of virulence factors, infection of mice, and survival inside cultured macrophages (Sittka *et al.*, 2007). Loss of Hfq function also abolishes *Salmonella* motility and deregulates more than 70 abundant proteins. This includes the accumulation of outer membrane proteins, which in turn cause chronic activation of the σ^E-mediated envelope stress response (Bang *et al.*, 2005; Bossi *et al.*, 2008; Figueroa-Bossi *et al.*, 2006; Sittka *et al.*, 2007, 2008). Given that Hfq primarily acts in concert with sRNAs, many of the above mentioned phenotypes in *Salmonella* may be attributed to loss of gene regulation by Hfq-associated sRNAs. Moreover, Hfq was implicated in the control of *Salmonella* gene expression changes induced by the low gravity condition experienced during spaceflight (Wilson *et al.*, 2007).

2.6.1.2. *Salmonella* sRNAs

Many of the *Salmonella* sRNAs were initially identified in *E. coli* (see Hershberg *et al.*, 2003; Papenfort *et al.*, 2008). However, the horizontally-aquired, *Salmonella* specific regions could encode novel sRNAs or other RNA elements that are absent in *E. coli*. Recently, two bioinformatics-based studies predicted several *Salmonella*-specific sRNAs (Padalon-Brauch *et al.*, 2008; Pfeiffer *et al.*, 2007). Pfeiffer *et al.*, 2007 searched for "orphan" pairs of σ^{70}-type promoters and Rho-independent transcription terminators in the intergenic regions (IGRs) of the *Salmonella* LT2 genome which led to the prediction of 46 sRNA candidates. These sRNA genes are absent from *E. coli* K12 but most of them have homologues in the relatively distant *Salmonella* species, *S. bongori*. This screen led to the discovery of the first sRNA from an enterobacterial pathogenicity island, *i. e.* the 80 nt InvR RNA that is expressed from the invasion gene locus SPI-1 (Pfeiffer *et al.*, 2007). The second study examined the genetic islands of *Salmonella* for the presence for putative sRNA genes, *i. e.* those IGRs that were $>$ 100 bps and showed $<$80% identity to their most similar sequence in *E. coli* K12 (Padalon-Brauch *et al.*, 2008). The predictions, which were also based upon seraches for orphan Rho-independent terminators, resulted in 28 sRNA candidate genes. Northern blot analysis confirmed expression of 19 island-encoded sRNAs, now denoted Isr (A, B, etc.), under a large panel of growth conditions reminiscent of the environments encountered by *Salmonella* upon host cell infection (Padalon-Brauch *et al.*, 2008). In addition, several validated sRNAs were shown to be differentially expressed upon *Salmonella* infection of macrophages. The function of these sRNAs is yet to be elucidated. A significant number of them overlap with the 5' or 3' ends of ORFs, and modulate the expression of these flanking ORFs or are in turn affected by those same genes. Overall, both studies predicted additional *Salmonella* sRNA candidates that remain to be verified experimentally.

2.6.2. *Helicobacter pylori*

The human pathogen *Helicobacter pylori* is the major cause of chronic superficial gastritis as well as peptic ulcer disease and furthermore leads to the development of adenocarcinoma and mucosa-associated lymphoid tissue (MALT) lymphoma of the stomach (reviewed in Blaser, 1998). Moreover, approximately 50 % of the world's population are infected with this microaerophilic and spiral-shaped, Gram-negative bacterium of the ϵ-proteobacter group. The *Helicobacter* species colonize the stomach of their hosts and are highly adapted to this acidic environment. Sequencing and annotation of the small and compact 1.67 Mb genomes of three *H. pylori* strains, namely, 26696, J99 and HPAG1, showed that *Helicobacter* features a very restricted repertoire of transcriptional regulators such as sigma factors, two-component systems and other response regulators (Alm *et al.*, 1999; Oh *et al.*, 2006; Tomb *et al.*, 1997). Specifically, *Helicobacter* contains only three sigma factors, RpoD (σ^{80}), FliA (σ^{28}), and RpoN (σ^{54}), only three two-component systems involved in transcriptional regulation, and two additional orphan response regulators. The ArsRS two-component system has been shown to be involved in gene-expression control during the acid-response (Pflock *et al.*, 2006). However, in order to cope with the various stresses it encounters during infection, *e. g.*, pH and nutrient fluctuations, *H. pylori* must have mechanisms to regulate the transcription of its genes.

2.6.2.1. Small regulatory RNAs in *H. pylori*

Several systematic screens have led to the identification of sRNAs in diverse bacteria (see Section 2.3). However, no regulatory sRNAs have so far been described in *Helicobacter* and none of the known enterobacterial sRNAs, except for the highly conserved housekeeping RNAs, tmRNA, RNase P RNA, SRP RNA, are conserved in *Helicobacter* on the primary sequence level (see for example `Rfam`[1] database). Moreover, biocomputational screens failed so far to find a 6S RNA (Section 2.1.3.1) homologue in the ϵ-subdivision of proteobacteria, a highly abundant RNA which could be identified in almost all groups of proteobacteria (Barrick *et al.*, 2005; Willkomm *et al.*, 2005). In *Helicobacter*, neither an Hfq nor an RNase E homologue have been described so far (Sun *et al.*, 2002; Tomb *et al.*, 1997). Thus, *Helicobacter* was often regarded as an example of a bacterium without any riboregulation. However, regulatory RNAs in *Helicobacter* might have other characteristics and functions than typical enterobacterial sRNAs and, thus, could involve different RNA-binding proteins from Hfq. Furthermore, sRNAs have been recently identified in *Streptococcus*, a bacterium which also lacks Hfq (Halfmann *et al.*, 2007). Similarly, although Hfq is absent from 10 of the 12 sequenced strains of the cyanobacterium *Prochlorococcus*, several sRNAs have been identified by comparative genome analysis in *Prochlorococcus* strain MED4 which also lacks Hfq (Axmann *et al.*, 2005). A recent study based on microarray expression profiling reported the presence of many additional sRNA genes as well as several antisense RNAs in the same strain (Steglich *et al.*, 2008). As a first hint that riboregulation is also present in *H. pylori*, Livny *et al.* (2006) predicted ≈ 50 sRNA candidates but without any experimental validation.

[1] `www.sanger.ac.uk/Software/Rfam/`

CHAPTER 3

MULTIPLE TARGETING OF ABC TRANSPORTER MRNAS BY GCVB SRNA

The interactions of numerous regulatory small RNAs (sRNAs) with single target mRNAs have been characterized extensively. In contrast, how sRNAs can regulate multiple, structurally unrelated mRNAs is less understood. The majority of the ≈ 20 enterobacterial sRNAs of known function are antisense RNAs that repress trans-encoded target mRNAs. Many sRNAs mask the ribosome binding sites (RBS) of their target, thus inhibiting ribosome entry on mRNA (see Section 2.1.1). Since the half-life of bacterial mRNAs is strongly affected by the association with ribosomes (Deana & Belasco, 2005), translation inhibition is often coupled to the decay of the repressed target, *e. g.*, by accelerating RNase E-mediated mRNA turnover (Section 2.2.2). Besides RNase E, the bacterial Sm-like protein Hfq has been identified as a key player in this type of translational silencing (Section 2.2.1).

Enterobacterial sRNAs have been shown to bind via short and/or imperfect base-pairing to their target mRNAs. For example, nine residues of RyhB sRNA interact with *sodB* mRNA and OxyS sRNA targets *fhlA* mRNA through the formation of two short kissing complexes of 9 and 7 bp, respectively (Section 2.1.1). Due to this limited sequence complementarity, the identification of new sRNA targets is difficult, and for many of the sRNAs studied to date, a single mRNA remains the only experimentally validated target. However, it was early recognized that some *E. coli* sRNAs, *e. g.*, DsrA and OxyS RNA, could regulate multiple genes and, in addition, recent biocomputational and experimental approaches predicted more sRNAs to target multiple mRNAs (see Section 2.5). For example, pulse expression of several *E. coli* and *Salmonella* sRNAs followed by global transcriptome profiling showed the rapid depletion of many mRNAs (Section 2.4.1.2). The kinetics of target decay observed in these experiments strongly argues that the regulated mRNAs are direct sRNA targets.

Multiple mRNA targeting by sRNAs could help bacteria to balance different transcriptional responses at the post-transcriptional level in response to stress or changes in nutrient availability. This additional layer of gene expression control could mediate the co-regulation of mRNAs that belong to different transcriptional regulons. However, only a few direct interactions of sRNAs with multiple mRNAs have been biochemically characterized in enterobacteria (Guillier & Gottesman, 2008; Lease *et al.*, 1998). Insights into the mechanistic aspects of multiple-target regulation have primarily come from investigations of RNAIII of the Gram-positive pathogen *Staphylococcus aureus* (Boisset *et al.*, 2007).

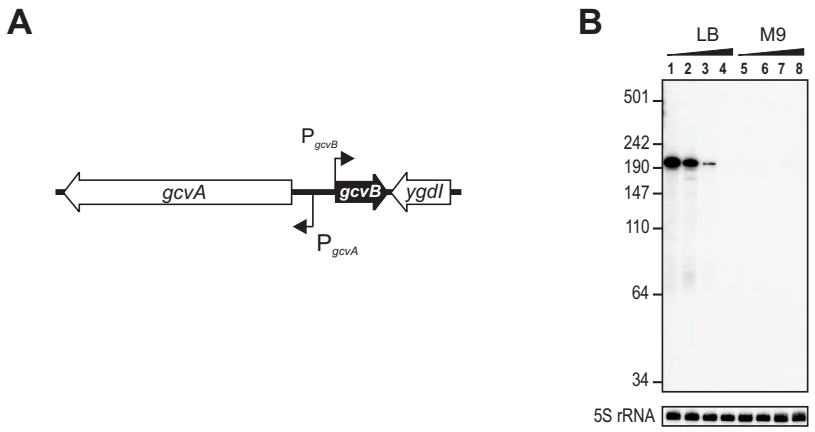


Figure 3.1: Genomic location and expression of *Salmonella* GcvB RNA. (A) The *Salmonella gcvB* gene is encoded in the *gcvA-ygdI* intergenic region in opposite orientation to these latter two genes. **(B)** Northern blot (autoradiogram) showing *gcvB* expression in nutrient-rich (L-broth) or nutrient-limiting (M9) media. Total RNA was prepared from *Salmonella* wild-type (strain JVS-0007) grown to exponential phase (OD_{600} of 0.4, lanes 1 and 5), early stationary phase (OD_{600} 1, lanes 2 and 6; OD_{600} 2, lanes 3 and 7), and late stationary phase cells (lanes 4 and 8). GcvB was detected with γ^{32}P-labelled oligonucleotide JVO-0749 complementary to the GcvB 5' region. Marker sizes are shown to the left. 5S rRNA probing (panel below) confirmed equal RNA loading.

In this Chapter, GcvB sRNA of *Salmonella enterica* serovar Typhimurium has been studied and turned out to directly act upon multiple mRNAs which commonly encode periplasmic substrate binding proteins of ABC uptake systems for amino acids and peptides. The *gcvB* gene was originally identified in *E. coli* and is controlled by GcvA/GcvR, the major transcription factors of the glycine cleavage system (Urbanowski *et al.*, 2000). A *gcvB* deletion caused constitutive synthesis of OppA and DppA, periplasmic binding proteins of the two major peptide transport systems (Abouhamad *et al.*, 1991; Higgins & Hardie, 1983), which are normally repressed in nutrient-rich growth conditions. Gene fusion experiments indicated that GcvB repressed *dppA* and *oppA* at the post-transcriptional level, yet the molecular mechanism remained elusive (Urbanowski *et al.*, 2000). Furthermore, the pleiotropic nature of *E. coli* and *Yersinia gcvB* mutants (McArthur *et al.*, 2006; Urbanowski *et al.*, 2000) and biocomputational predictions (Tjaden *et al.*, 2006) suggested that GcvB may have additional mRNA targets.

This section provides biochemical and genetic evidence that a conserved G/U-rich region within GcvB mediates translational repression of seven ABC transporter mRNAs. This G/U-rich element, which was revealed by alignment of GcvB homologues of distantly related bacteria, is strictly required for GcvB target recognition. Analysis of target gene fusion regulation *in vivo*, as well as *in vitro* structure probing and translation assays showed that GcvB represses its target mRNAs by binding to extended C/A-rich regions which may also serve as translational enhancer elements. In some cases (*oppA*, *dppA*), GcvB repression can be explained by masking the ribosome binding site (RBS) to prevent 30S subunit binding. However, GcvB can also effectively repress translation by binding to target mRNAs at upstream sites, outside the RBS. Specifically, GcvB represses *gltI* mRNA translation at a C/A-rich target site located at positions -57 to -45 relative of the start codon.

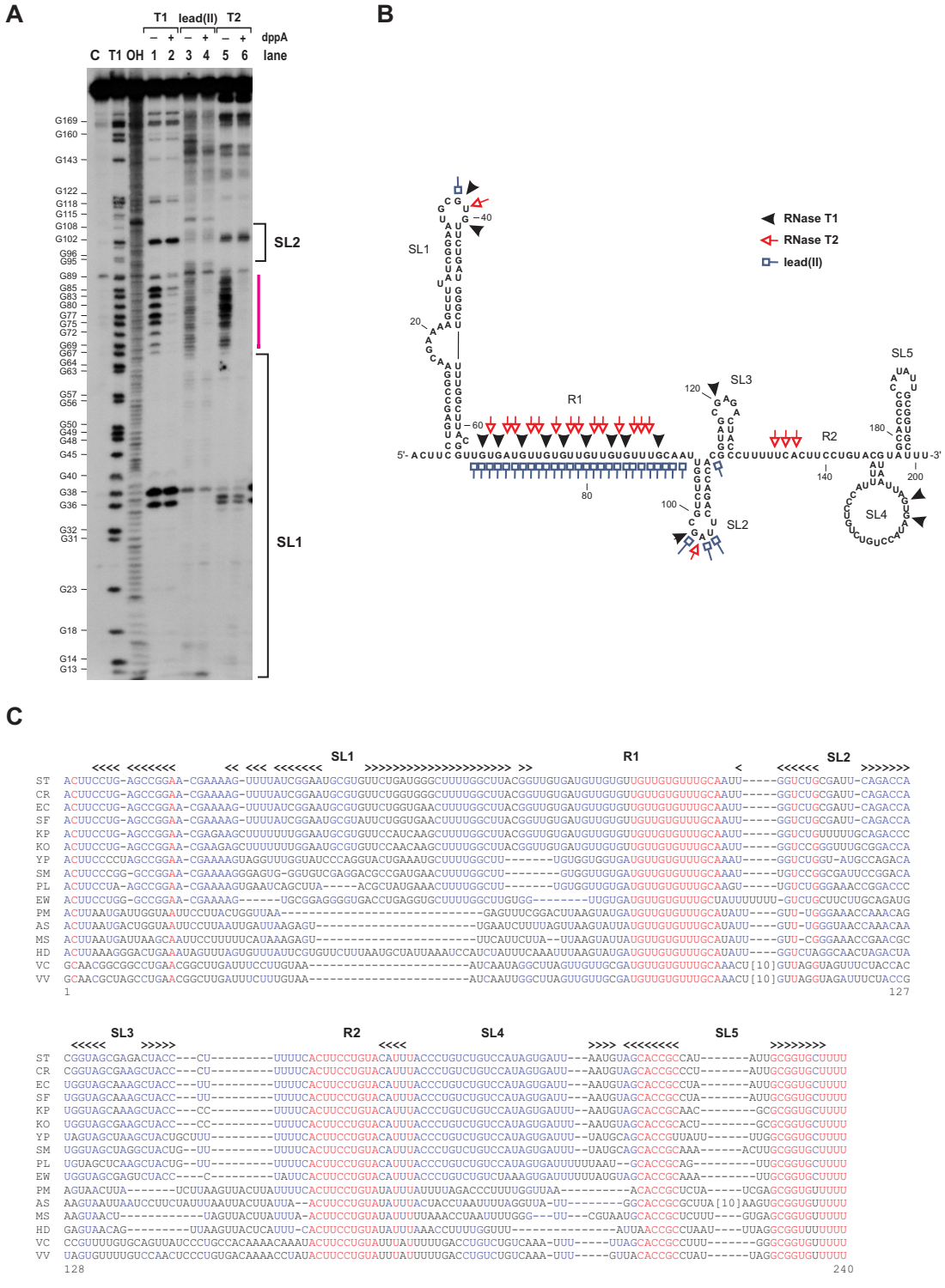
This latter mode of action may also apply to other regulatory sRNAs. Taken together, this suggests highly conserved regions in sRNAs, and mRNA regions distant from Shine-Dalgarno (SD) sequences as important elements for the identification of sRNA targets. Most of the work described in this section has been published in Sharma *et al.*, 2007.

3.1. Results

3.1.1. Characterization of GcvB sRNA in *Salmonella*

The *Salmonella* and *E. coli gcvB* genes are ≈95% identical in the GcvB RNA region (Argaman *et al.*, 2001; Urbanowski *et al.*, 2000) and both are encoded in the intergenic region between *gcvA*, a transcriptional activator of the glycine cleavage system, and *ygdI*, a protein of unknown function (Figure 3.1A). Expression of the ≈200 nt *Salmonella* GcvB RNA was confirmed on Northern blots of RNA extracted from different growth phases and media. GcvB is expressed in exponentially growing *Salmonella* in rich medium (L-broth), but is not detectable in stationary phase or upon growth in minimal medium (Figure 3.1B). This pattern is reminiscent of *E. coli* GcvB (Argaman *et al.*, 2001), and in keeping with the postulated GcvB repressor function of peptide transporters under nutrient-rich conditions (Urbanowski *et al.*, 2000). GcvB RNA folds into a rather unusual secondary structure (Fig. 3.2A and B). *In vitro* structure probing revealed the presence of several stem-loop structures which are separated by two single-stranded linker regions, R1 and R2 (see Figure 3.2). Especially the linker region R1 is interesting because it consists almost exclusively of G/U residues. As single stranded regions have been shown to be involved in sRNA/target interactions, these linker regions could be involved in target recognition by GcvB RNA.

Figure 3.2 *(facing page)*: Structure mapping of 5'end labelled GcvB RNA and proposed secondary structure. **(A)** 5'end-labelled GcvB RNA (≈5 nM) was subjected to RNase T1, lead(II), and RNase T2 cleavage in absence (lanes 1, 3, 5) and presence of different concentrations of cold *dppA* leader (lanes 2, 4, 6: ≈250 nM final concentration). Lane C: untreated GcvB RNA. Lane T1: RNase T1 ladder of GcvB under denaturing conditions. The position of cleaved G residues is given at the left of the gel. Lane OH: Alkaline ladder. The approximate positions of stem-loop structures SL1 and SL2 according to the GcvB RNA structure shown in **(B)** are indicated by vertical bars to the right of the gel. The region protected in GcvB RNA by *dppA* leader is marked by a magenta bar. **(B)** Proposed secondary structure of GcvB based on *in vitro* structure mapping. Cleavages according to **(A)** by RNAse T1, T2, or lead(II) are indicated by black, red, and blue arrows, respectively. Due to low resolution in the gel part that corresponds to the 3' part of GcvB RNA, no detailed cleavages are indicated in this region. **(C)** Alignment of a representative subset of GcvB RNAs identified by computer-based searches (see Figure 10.1 in the Appendix). Numbering of residues and the positions of stem-loop structures (SL1 to SL5; indicated by arrowheads) follows the *Salmonella* GcvB RNA sequence in **(B)**. The conserved regions, R1 and R2, common to all known GcvB sequences are indicated. (ST) *Salmonella typhimurium*; (CR) *Citrobacter rodentium*; (EC) *Escherichia coli* K12; (SF) *Shigella flexneri*; (KP) *Klebsiella pneumoniae*; (KO) *Klebsiella oxytoca*; (YP) *Yersinia pestis*; (SM) *Serratia marcescens*; (PL) *Photorhabdus luminescens*; (EW) *Erwinia carotovora*; (PM) *Pasteurella multocida*; (AS) *Actinobacillus succinogenes*; (MS) *Mannheimia succiniciproducens*; (HD) *Haemophilus ducreyi*; (VC) *Vibrio cholerae*; (VV) *Vibrio vulnificus*.


A
T1 lead(II) T2
dppA
C T1 OH 1 2 3 4 5 6 lane
SL2
SL1
B
RNase T1
RNase T2
lead(II)
SL1
SL2
SL3
SL4
SL5
R1
R2
5'- ACUUC GUUGUGAUGUUGUGUUGUUGUGUUUGCAAU CGCCUUUUUCACUUCCUGUACGUAUUU-3'
C
SL1
R1
SL2
SL3
R2
SL4
SL5

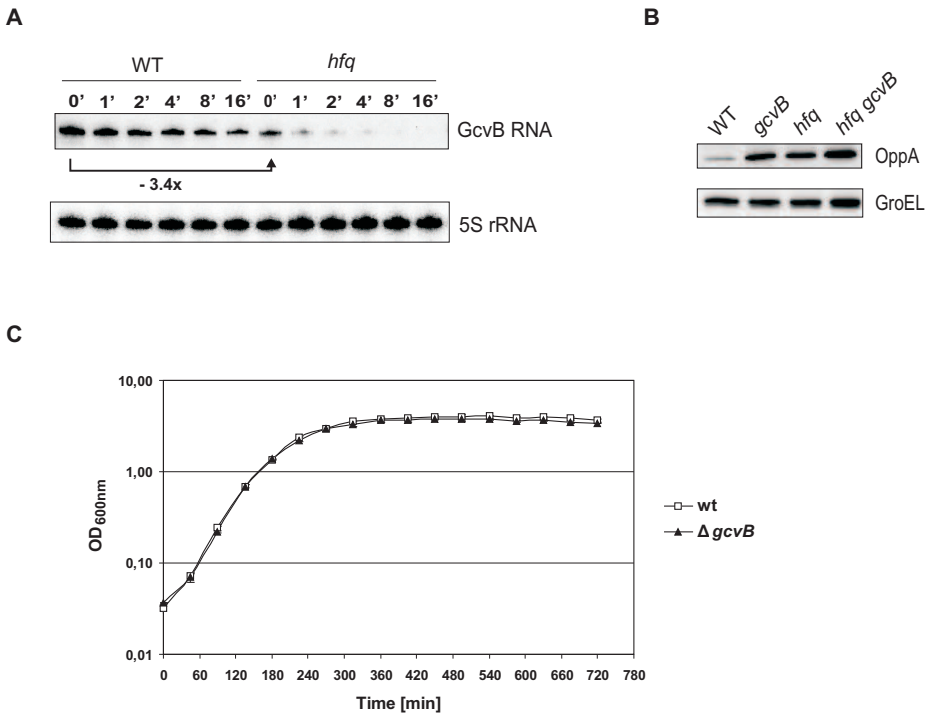


Figure 3.3: GcvB stability in *Salmonella* wild-type and Δ*hfq* and growth curves. (A) Cultures of *Salmonella* wt and a *hfq* deletion strain were grown to an OD_{600} of 0.4, and RNA stabilities were determined by rifampicin treatment (final concentration of 500 μg/ml). 5 ml cells were withdrawn prior to or 1, 2, 4, 8, and 16 min after rifampicin addition, mixed with 0.2 vol of stop solution (5% water-saturated phenol, 95% ethanol), and snap-frozen in liquid nitrogen. Upon thawing, RNA was isolated and analysed by Northern hybridization. Quantification of the blots revealed a 3.4-fold reduction in GcvB RNA levels in the *hfq* deletion strain. The GcvB half-life ($\approx$3 min in wild type) is decreased to $\approx$30 sec in the *hfq* deletion strain. **(B)** Both GcvB and Hfq act to repress OppA protein synthesis. Protein samples of *Salmonella* wild-type, *hfq* or *gcvB* single deletion, and the *hfq gcvB* double deletion strain (JVS-0007, -0255, -0236, and -0617, respectively) grown to exponential phase (OD_{600} of 0.4) were subjected to Western blot analysis with OppA (upper panel) or GroEL antisera (lower panel; loading control). Quantification of the OppA signals, followed by normalization to GroEL levels, revealed a 6.1-fold, 5.4-fold or 8.3-fold increase upon *gcvB*, *hfq*, or *gcvB hfq* double deletion, respectively. **(C)** OD_{600} values of triplicate cultures of *Salmonella* wild-type (JVS-0007) or *gcvB* deletion strain (JVS-0236) grown in LB medium were determined in 45 min intervals. The curves show the OD_{600} average values and standard deviations that were calculated from the triplicates (open squares: wild-type; filled triangles: Δ*gcvB*). No growth defect was observed for the *Salmonella* Δ*gcvB* strain compared to the wild-type strain.

Several observations suggested that GcvB is an Hfq-dependent sRNA. GcvB co-immunoprecipitates with Hfq in extracts of *E. coli* (Zhang *et al.*, 2003) and *Salmonella* (Sittka *et al.*, 2008), and is unstable in Δ*hfq* strains of *E. coli* (Urban & Vogel, 2007) and *Salmonella* (Figure 3.3A). Since repression of the predicted GcvB target, *oppA*, is abrogated in *hfq* mutants of both *E. coli* and *Salmonella* (Sittka *et al.*, 2007; Ziolkowska *et al.*, 2006), OppA protein levels were compared in *Salmonella* strains deleted for *gcvB*, *hfq* or both (Fig. 3.3B).

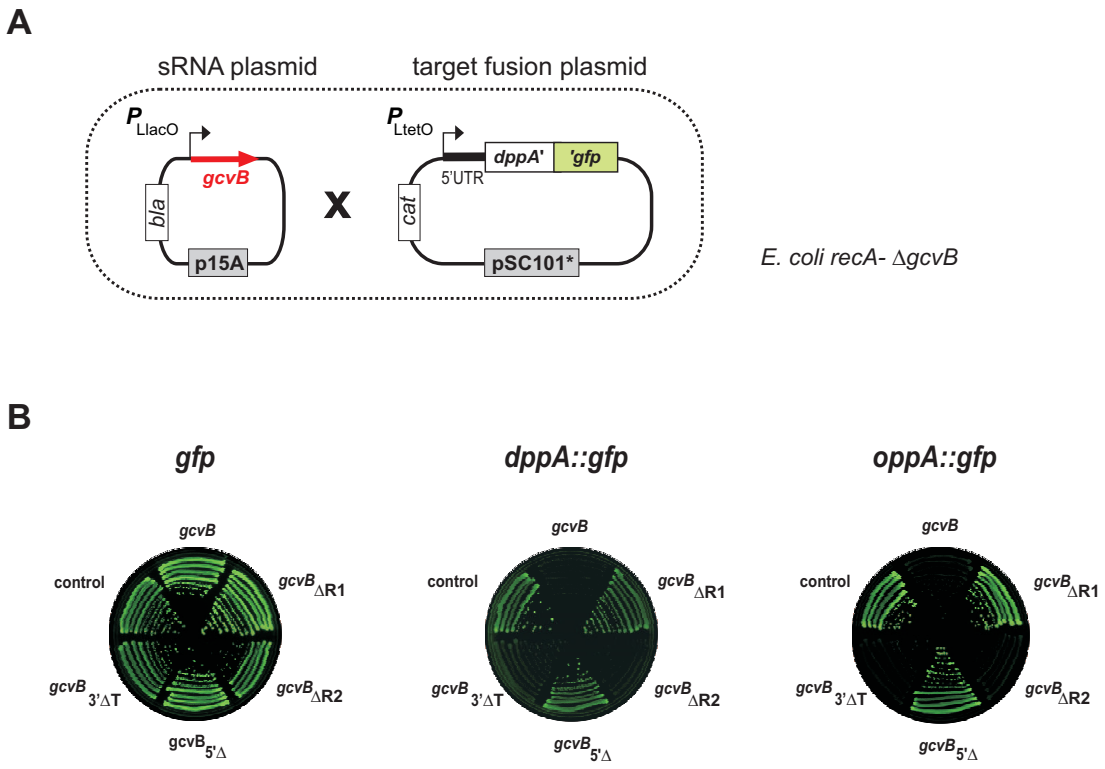


Figure 3.4: Regulation of *Salmonella oppA::gfp* and *dppA::gfp* fusions by GcvB wild-type and mutant RNAs. (A) GcvB and translational fusions of *oppA* or *dppA* to *gfp* are constitutively expressed from compatible plasmids in an *E. coli* Δ*gcvB recA*$^-$ strain (Urban & Vogel, 2007). The fusions include the entire 5' UTR (determined by 5'RACE; *dppA*, 163 bp; *oppA*, 162 bp) and result in in-frame fusions of the 10th (*dppA*) or 17th (*oppA*) codon to the amino terminus of GFP. **(B)** Agar plate-based assay of colony fluorescence of *E. coli* strains carrying control plasmid pXG-1 (*left*; expresses full-length *gfp*), the *dppA::gfp* (*middle*; plasmid pJL18-1) or the *oppA::gfp* (*right*; pJL19-1) fusion plasmid, each in combination with control vector pTP11, or plasmids expressing *Salmonella* wild-type GcvB RNA (pP_L*gcvB*) or four of the mutant alleles (plasmids pP_L*gcvB*$_{\Delta R1}$, pP_L*gcvB*$_{\Delta R2}$, pP_L*gcvB*$_{5'\Delta}$, p*gcvB*$_{3'\Delta T}$) shown in Figure 3.8A. All strains displayed normal colony morphology (not shown). Reduced colony fluorescence of the *dppA::gfp* or *oppA::gfp* fusions strains upon sRNA co-expression indicates regulation at the post-transcriptional level. *(Left)* GcvB has no effect on the expression of *gfp* alone, confirming that repression is specific to the cloned *dppA* and *oppA* regions (middle and right).

Deletion of *gcvB* did not impair *Salmonella* growth in rich broth (Fig. 3.3C), similar to what was reported for *E. coli* (Urbanowski *et al.*, 2000). The single *gcvB* or *hfq* mutations each elevated OppA levels, whereas the double mutation had no pronounced cumulative effect (Fig. 3.3B). This predicts GcvB and Hfq to act in concert, and GcvB to be an Hfq-dependent antisense RNA that regulates *trans*-encoded mRNA(s).

3.1.2. GcvB targets *dppA* and *oppA* mRNAs *in vivo* and *in vitro*

To study *dppA* and *oppA* mRNA repression by GcvB, translational fusions to the amino terminus of green fluorescent protein (GFP) were constructed (Urban & Vogel, 2007). The *Salmonella dppA* or

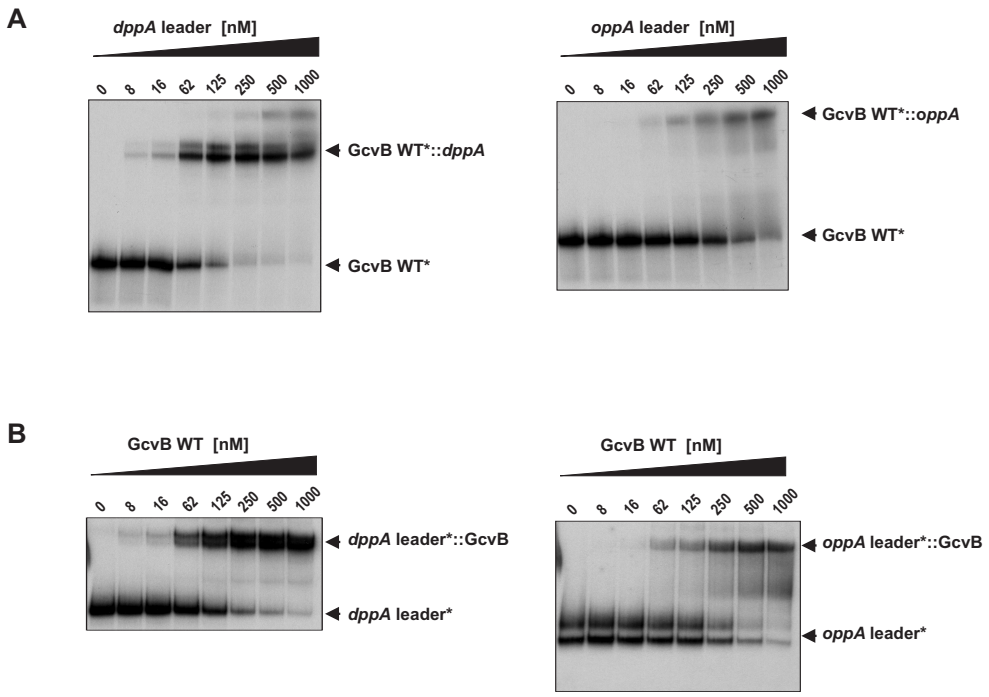


Figure 3.5: Gel-mobility shift assays of GcvB WT RNA and *dppA* and *oppA* leader. (A) *In-vitro* binding of GcvB RNA and *dppA/oppA* leader was performed as described in Section 8.3.7 in Material and Methods. Approximately 5 nM of γ^{32}P-labelled GcvB RNA was incubated with increasing concentrations of unlabelled *dppA* or *oppA* leader (final concentrations in nM above the lanes). Following 15 min incubation at 37°C, samples were run on a native 6% PAA gel. Shown is an autoradiograph of the gel. **(B)** Same experimental procedure as above but with 5'-labelled *dppA/oppA* leader and increasing concentrations of unlabelled GcvB RNA.

oppA 5' regions spanning the entire 5' UTR and 30 bp (*dppA*) or 50 bp (*oppA*) of the coding region were cloned into the low-copy *gfp* fusion vector, pXG10 (Figure 3.4A). The cloning strategy used here ensures that the fusions are transcribed from the native +1 site of *dppA* or *oppA* without adding additional sequences at the 5' end. Transcription is driven by a constitutive $P_{LtetO-1}$ promoter (Lutz & Bujard, 1997) to specifically assay post-transcriptional regulation. The *Salmonella gcvB* gene was placed under control of a constitutive $P_{LlacO-1}$ promoter on a compatible mid-copy plasmid (Lutz & Bujard, 1997) resulting in plasmid pP_L *gcvB* (Figure 3.4A).

The fusion plasmids as well as control plasmid pXG-1 expressing full-length GFP were each combined with either pP_L *gcvB* or control vector pTP11 in an *E. coli* Δ*gcvB* strain. The specific repression of *dppA::gfp* and *oppA::gfp* by pP_L *gcvB* was evident from strongly reduced colony fluorescence of these strains on agar plates (Figure 3.4 B) which established that GcvB regulates *dppA* and *oppA* in the 5' mRNA region. To map the GcvB interaction sites, RNAs of the previously cloned *dppA* and *oppA* fragments were synthesized *in vitro* and structural probing experiments were performed. Gel mobility shift assays showed that GcvB formed singular complexes with either of the two RNAs under standard *in vitro* conditions (Fig. 3.5). RNA structure probing of these complexes

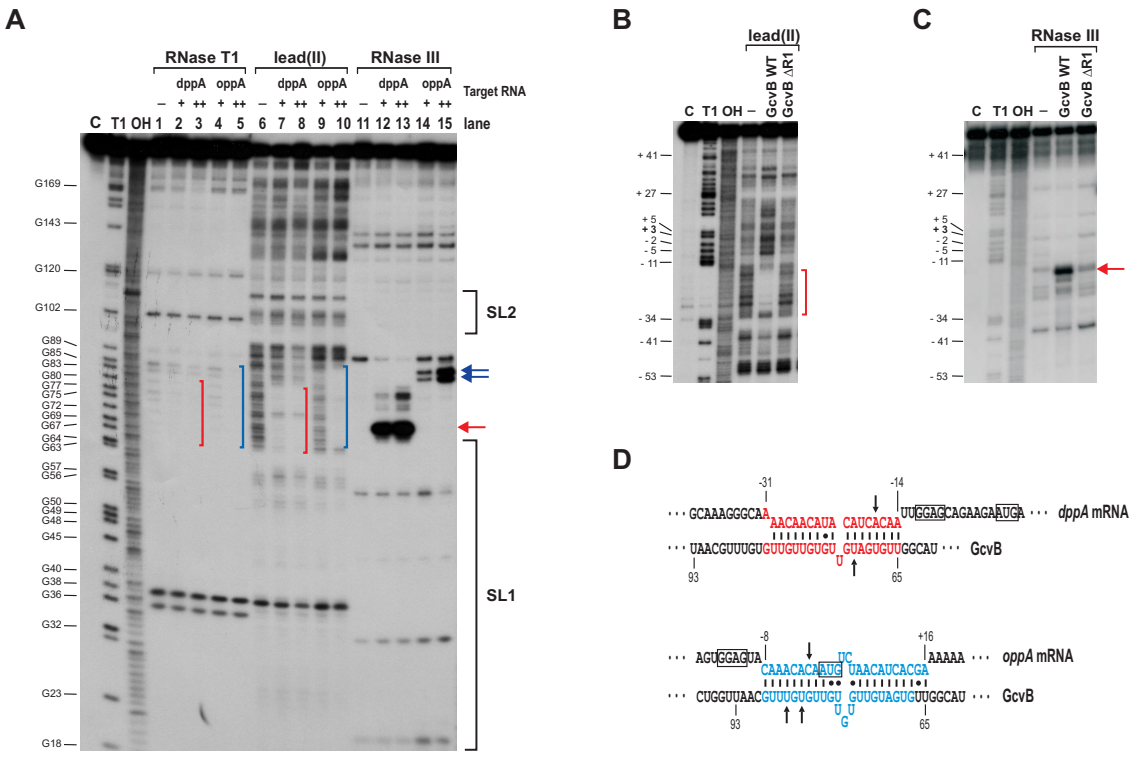


Figure 3.6: Identification of GcvB binding sites on *oppA* and *dppA* mRNAs by *in vitro* probing. (A) 5'end-labelled GcvB RNA (≈5 nM) was subjected to RNase T1, lead(II), and RNase III cleavage in the absence (lanes 1, 6, 11) and presence of different concentrations of cold *dppA* or *oppA* RNAs (final concentration in lanes 2, 4, 7, 9, 12, 14: ≈50 nM; lanes 3, 5, 8, 10, 13, 15: ≈500 nM). The synthesized target RNA fragments comprise regions -163/+72 (*dppA*) and -162/+57 (*oppA*) relative to the AUG start codon. (Lane C) Untreated GcvB RNA. (Lane T1) RNase T1 ladder of hydrolysed denatured GcvB RNA. The position of cleaved G residues is given left of the gel. (Lane OH) Alkaline ladder. Vertical bars indicate the GcvB region protected by *dppA* RNA (red) and *oppA* RNA (blue). Arrows denote specific RNase III cleavage of GcvB in the presence of *dppA* and *oppA* RNAs. The approximate positions of stem-loop structures SL1 and SL2 according to the GcvB RNA structure shown in Figure 3.2 B are indicated to the right of the gel. **(B)** 5' end-labelled *dppA* RNA treated with lead(II) **(B)** or RNase III **(C)** in the absence (lane -) or presence of GcvB wild-type or GcvB$_{\Delta R1}$ mutant RNA (lacks residues 66-89) as indicated above the gels. GcvB but not ΔR1 mutant RNA protects a ≈20 nt region (vertical red bar in **B**) in *dppA* from lead(II) cleavage, and only GcvB WT induces strong RNase III cleavage of the *dppA* RNA (red arrow in **C**). Residue G "+3" set in bold is the G in the *dppA* AUG start codon. **(D)** Predicted RNA duplexes formed by GcvB with the *dppA* or *oppA* mRNAs. Vertical arrows denote RNase III cleavage sites. SD and AUG start codon sequences are boxed. The coloured residues were protected from lead(II) cleavage upon duplex formation (see **A**-**C** and Figure 3.7).

with RNase T1 and lead(II) acetate showed that increasing concentrations of *dppA* or *oppA* RNAs resulted in "footprints" on 5' end-labelled GcvB RNA (Fig. 3.6A), which were most pronounced with lead(II) probing (lanes 6-10). The *dppA* and *oppA* RNAs each protect ≥19 residues within the single-stranded, highly G/U-rich region between stem-loops SL1 and SL2 of GcvB (compare Figs. 3.6A and 3.2A & B).

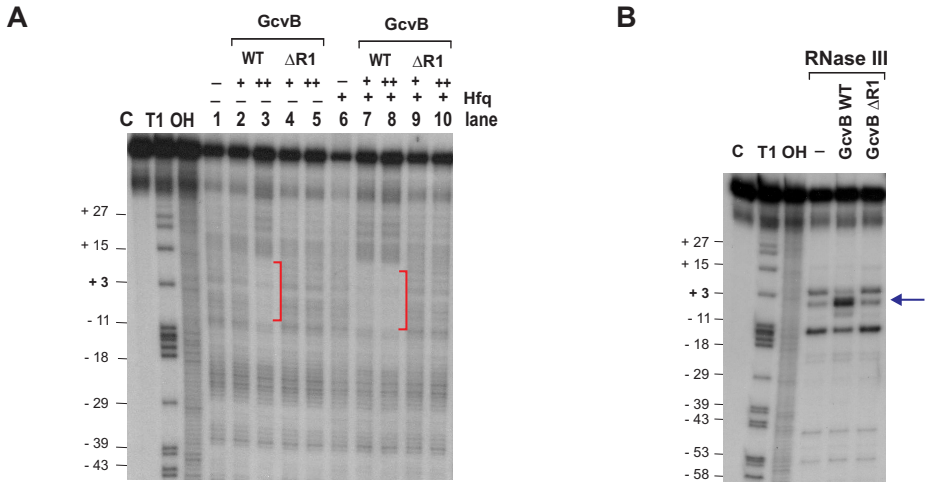


Figure 3.7: *In-vitro* footprinting and RNase III cleavage of *oppA*-leader/GcvB RNA complexes. 5' end-labelled *oppA* leader RNA was treated with lead(II) **(A)** or RNase III **(B)** in the absence (lanes 1 and 6) or presence of different concentrations of cold GcvB wild-type (lanes 2, 3, 7, and 8) or its ΔR1 mutant RNA (residues 66 - 89 deleted, lanes 4, 5, 9, and 10) as indicated above the gels. Furthermore, lead(II) cleavage was carried out in the absence (lanes 1 - 5) and presence (lanes 6 -10) of Hfq (50 nM final). GcvB WT but not ΔR1 protects a ≈ 20 nt region (vertical red bar) in *oppA* from lead(II) cleavage, and only GcvB WT induces strong RNase III cleavage of the *oppA* RNA (blue arrow). Hfq clearly improves binding of GcvB RNA to *oppA* leader, since the footprint for the lower GcvB concentration is only visible in the presence of Hfq (compare lanes 2 and 7). Positions of cleaved G residues in the *oppA* sequence are given as distance to the *oppA* start codon, *e. g.*, position "+ 3" which is indicated in bold corresponds to the G residue of the *oppA* start codon AUG.

The reciprocal experiment, *i. e.* probing of labelled *dppA* or *oppA* RNAs in complex with GcvB, identified the GcvB binding site on the mRNAs. Wild-type GcvB RNA yields a strong footprint on the *dppA* (Fig. 3.6B) and *oppA* (Fig. 3.7) RNAs, which correspond to regions -31 to -14, and -8 to +16, respectively, relative to the AUG start codon (Fig. 3.6D). In contrast, no footprints were obtained with a mutant RNA, $GcvB_{\Delta R1}$, which lacks the G/U-rich region (Figs. 3.6B and 3.7).

The duplexes shown in Figure 3.6D were further supported by cleavage of the GcvB/*dppA* and GcvB/*oppA* complexes with the double strand-specific nuclease RNase III. While several weak RNase III-dependent bands were observed in GcvB RNA alone (Fig. 3.6A, lane 11), the enzyme cleaved GcvB strongly and specifically in the G/U-rich region in the presence of either *dppA* or *oppA* RNA (lanes 12-15). Reciprocally, the *dppA* and *oppA* RNAs were specifically cleaved in the proposed GcvB binding site (Figs. 3.6C and 3.7B). Collectively, these results indicated the single-stranded, G/U-rich region of GcvB as an important determinant for target recognition.

3.1.3. A conserved G/U-rich region mediates GcvB repression *in vivo*

The importance of the G/U-rich element was also evident from its strong conservation in GcvB sequences of distantly related species. Computer-based searches predicted *gcvB* genes in many enterobacteria as well as in *Pasteurellaceae* and *Vibrionaceae* (Fig. 10.1 in the Appendix). The

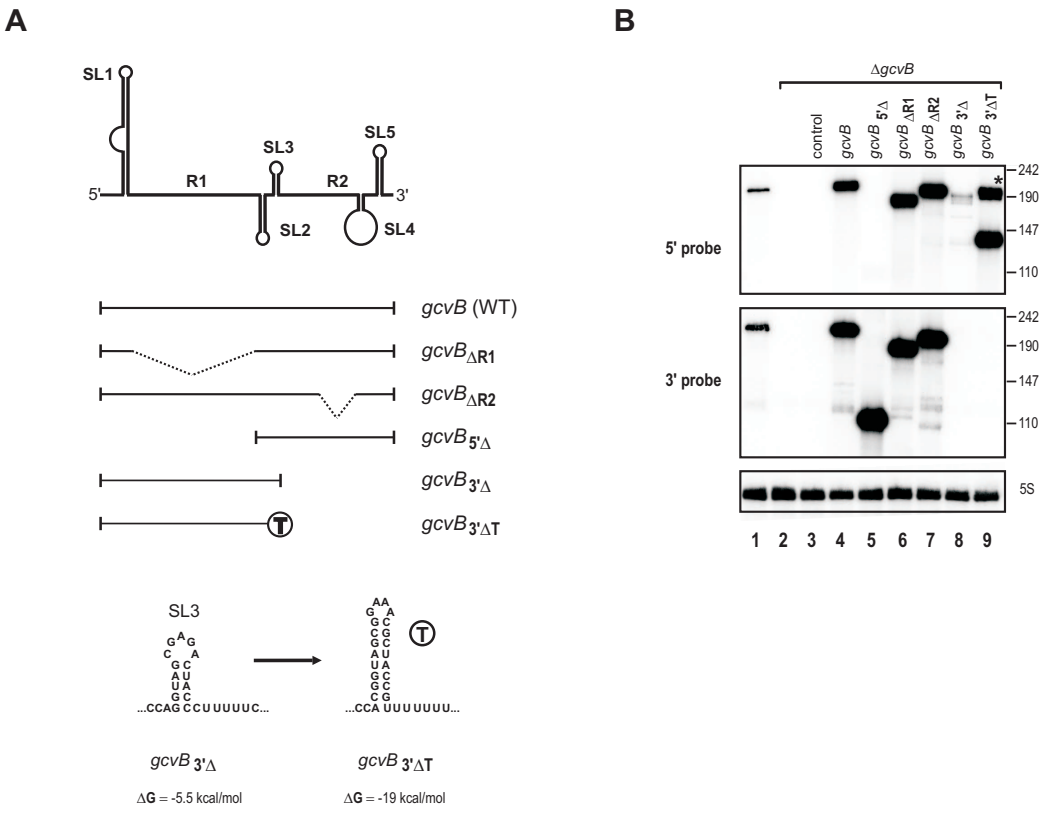


Figure 3.8: Mutant alleles of *gcvB* and their expression. **(A)** Horizontal bars below a schematic drawing of GcvB RNA denote the *gcvB* fragments expressed from mutant alleles; dotted lines denote internal deletions. Alleles $gcvB_{\Delta R1}$ and $gcvB_{\Delta R2}$ lack residues 66-89 (deletion of the G/U-rich R1 sequence) or 136-144 (deletion of R2), respectively. The truncated $gcvB_{5'\Delta}$ and $gcvB_{3'\Delta}$ alleles lack residues 1-91 (SL1 and R1) or 135-201 (R2 to SL5), respectively. In allele $gcvB_{3'\Delta T}$, which derives from $gcvB_{3'\Delta}$, SL3 was modified towards a transcription terminator as shown below. **(B)** Northern blots showing expression of *gcvB* or mutant alleles when borne on a mid-copy plasmid under control of the *gcvB* promoter in *Salmonella* (lanes 4-9). RNA was isolated from *Salmonella* grown to an OD_{600} of 1, and except lane 1 (wild type; JVS-0007) from a ΔgcvB genetic background (JVS-0236). The strains in lanes 3 to 9 carried control vector, pTP11, or plasmids p*gcvB*, p$gcvB_{5'\Delta}$, p$gcvB_{\Delta R1}$, p$gcvB_{\Delta R2}$, p$gcvB_{3'\Delta}$, p$gcvB_{3'\Delta T}$, respectively. Marker sizes are shown to the right. The upper blot was probed with labelled oligo JVO-0749, which is complementary to the GcvB 5' region (20-4 bp); the lower blot with JVO-0750 complementary to bp 172-150. The asterisk denotes transcriptional read-through to the *rrnB* terminator located downstream on the plasmids.

location of these candidate genes adjacent to and divergent from *gcvA*, and the conservation of promoter and terminator elements argue that these are *gcvB* homologues. Although the corresponding RNA sequences proved of enormous sequence diversity, it was obvious that much of the G/U-rich linker between stem-loops SL1 and SL2 is highly conserved in all identified *gcvB* sequences (Figure 3.2C). This linker region will be referred to as R1 (Conserved **R**egion **1**). The single-stranded region between SL3 and SL4 contains another highly conserved sequence (ACUUCCUGUA) found immediately upstream of SL4; this sequence will be referred to as R2.

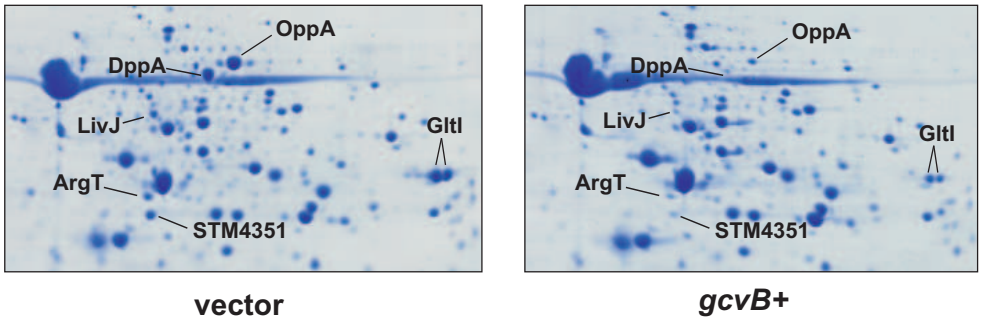


Figure 3.9: 2D analysis of periplasmic proteins. *Salmonella* ΔgcvB carrying control plasmid pTP11 (*left*) or *gcvB* mid-copy plasmid pTP05 (*right*), were grown to early stationary phase, and periplasmic protein fractions of these bacteria were resolved on 2D gels. Relevant subsections of the gels are shown. Protein spots indicating repression by GcvB are labelled.

To assess the role of these two regions in target regulation, mutant alleles $gcvB_{\Delta R1}$, $gcvB_{\Delta R2}$, and $gcvB_{5'\Delta}$, the latter being a 5' truncated *gcvB* lacking SL1 and R1, were constructed (Figure 3.8A). Plasmids carrying these alleles expressed distinct GcvB-derived RNAs at levels similar to wild-type *gcvB* (Fig. 3.8B). Next, these alleles were cloned under control of the $P_{LtetO-1}$ promoter to test their ability to regulate the *dppA::gfp* and *oppA::gfp* fusions. Figure 3.4B shows that loss of R1 ($gcvB_{\Delta R1}$ and $gcvB_{5'\Delta}$) abrogated *dppA* and *oppA* fusion repression, whereas loss of R2 ($gcvB_{\Delta R2}$) had no effect.

Is the 5' part of GcvB (including R1) sufficient to confer target repression *in vivo*? Urbanowski *et al.* (2000) postulated that SL3 of GcvB, which is followed by a U-stretch in many GcvB species (Fig. 3.2C), may serve as a ρ-independent transcription terminator, leading to the expression of a shorter GcvB RNA. However, truncation of *gcvB* after residue T_{134} (allele $gcvB_{3'\Delta}$, Fig. 3.8A, expected to yield a ≈134 nt GcvB RNA) did not result in such an RNA (Fig. 3.8B), whereas modification of SL3 created a functional terminator (Fig. 3.8A; allele $gcvB_{3'\Delta T}$) and led to detection of a ≈130 nt RNA (Fig. 3.8B). Figure 3.4B shows that the 3' truncated GcvB RNA of $gcvB_{3'\Delta T}$ was fully active in *dppA* and *oppA* fusion repression. Taken together, these *in vivo* experiments further support a key role of the G/U-rich element, R1, of GcvB for *dppA* and *oppA* mRNA regulation.

3.1.4. More GcvB targets

The observation that additional periplasmic proteins accumulated in *gcvB* or *hfq* mutant strains (Sittka *et al.*, 2007; Urbanowski *et al.*, 2000) prompted to screen for more GcvB targets. Analysis of periplasmic proteins from a *Salmonella gcvB* overexpression strain predicted *gltI*, *livJ*, *argT*, and STM4351 as further candidate targets (Fig. 3.9). In an independent approach, the `RNAhybrid` algorithm (Rehmsmeier *et al.*, 2004) was used to search for stable RNA duplexes of the R1 sequence of GcvB with the 5' regions of all *Salmonella* genes. This biocomputational search supported *gltI*, *livJ*, *argT*, and STM4351, and further predicted *livK* as a GcvB target (Fig. 3.10A). Intriguingly, the predicted GcvB binding sites on these mRNAs do not match in sequence but are rich in C and A residues, which also holds true for *dppA* and *oppA* (Fig. 3.6D).

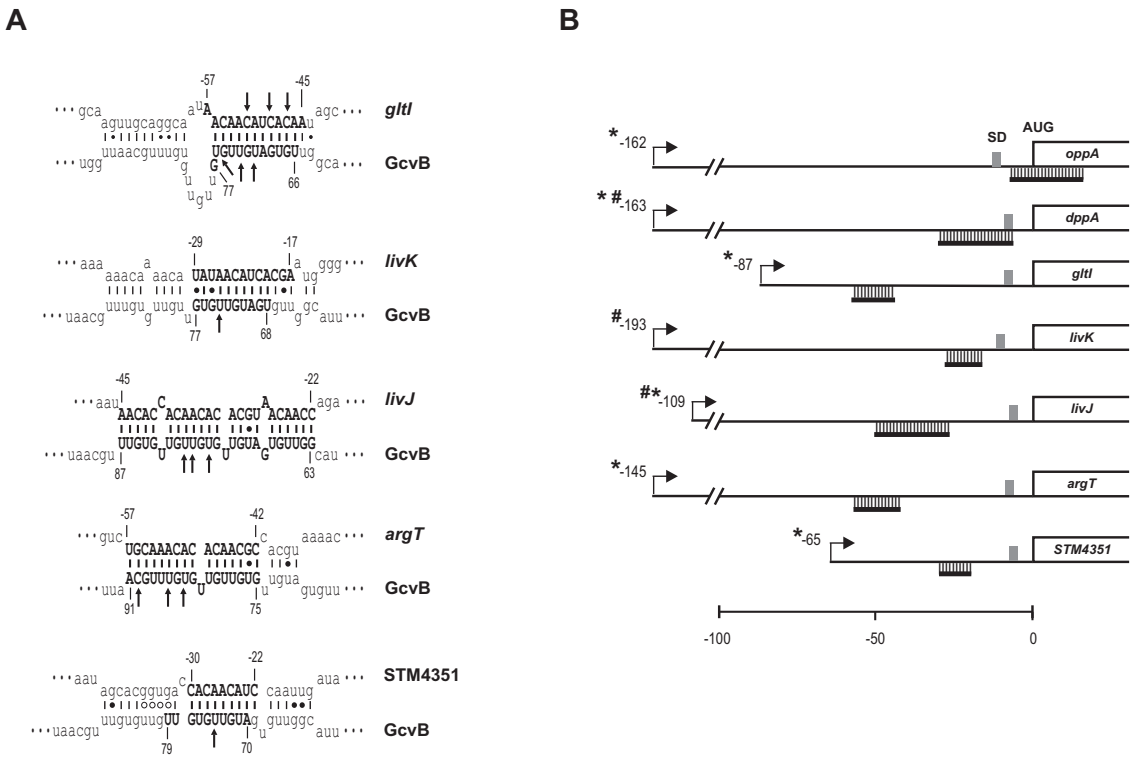


Figure 3.10: Summary of *in vitro* probing results of GcvB-target complexes. (A) Proposed RNA duplexes formed by GcvB with five periplasmic transporter mRNAs (*gltI*, *livK*, *livJ*, *argT*, and STM4315). Positions in the target sequences are given as distance to the start codon. Vertical arrows denote *in vitro* RNase III cleavage of the GcvB-*gltI* complex, and of GcvB in complex with the four other targets (Figs. 3.11A and 3.12B). Residues that showed protection in *in vitro* footprinting experiments (Figs. 3.11 and 3.12A) are set in bold and capitalized. Biocomputational prediction of target sites proposed the formation of extended duplexes around the interaction sites mapped by footprint analysis. Dark circles indicate G:U wobble base-pairs; open circles indicate more than three consecutive G:U pairs. **(B)** Location of GcvB binding sites on target mRNAs. 5' UTRs of target genes are drawn to scale. Asterisks indicate promoter positions that were mapped by 5'RACE; '#' indicates promoters according to EcoCyc (`www.ecocyc.org`) annotations. SD sequences are shadowed. The GcvB binding sites on the mRNAs are indicated by horizontal bars.

The predicted binding sites were subsequently confirmed by *in vitro* probing experiments as above (shown in Figure 3.11; summarized in Figure 3.10). Although the footprint obtained for *gltI* mRNA is relatively weak, the predicted interaction is strongly supported by specific RNase III cleavage in the presence of GcvB (Fig. 3.11A). Conversely, R1 is the GcvB region that is most strongly protected upon incubation with the target RNA fragments (Fig. 3.12). When in complex with target RNAs, GcvB was cleaved by RNase III exclusively in the R1 sequence, except with a *livK*-derived fragment which also promotes cleavage in the GcvB 3' part (Fig. 3.12).

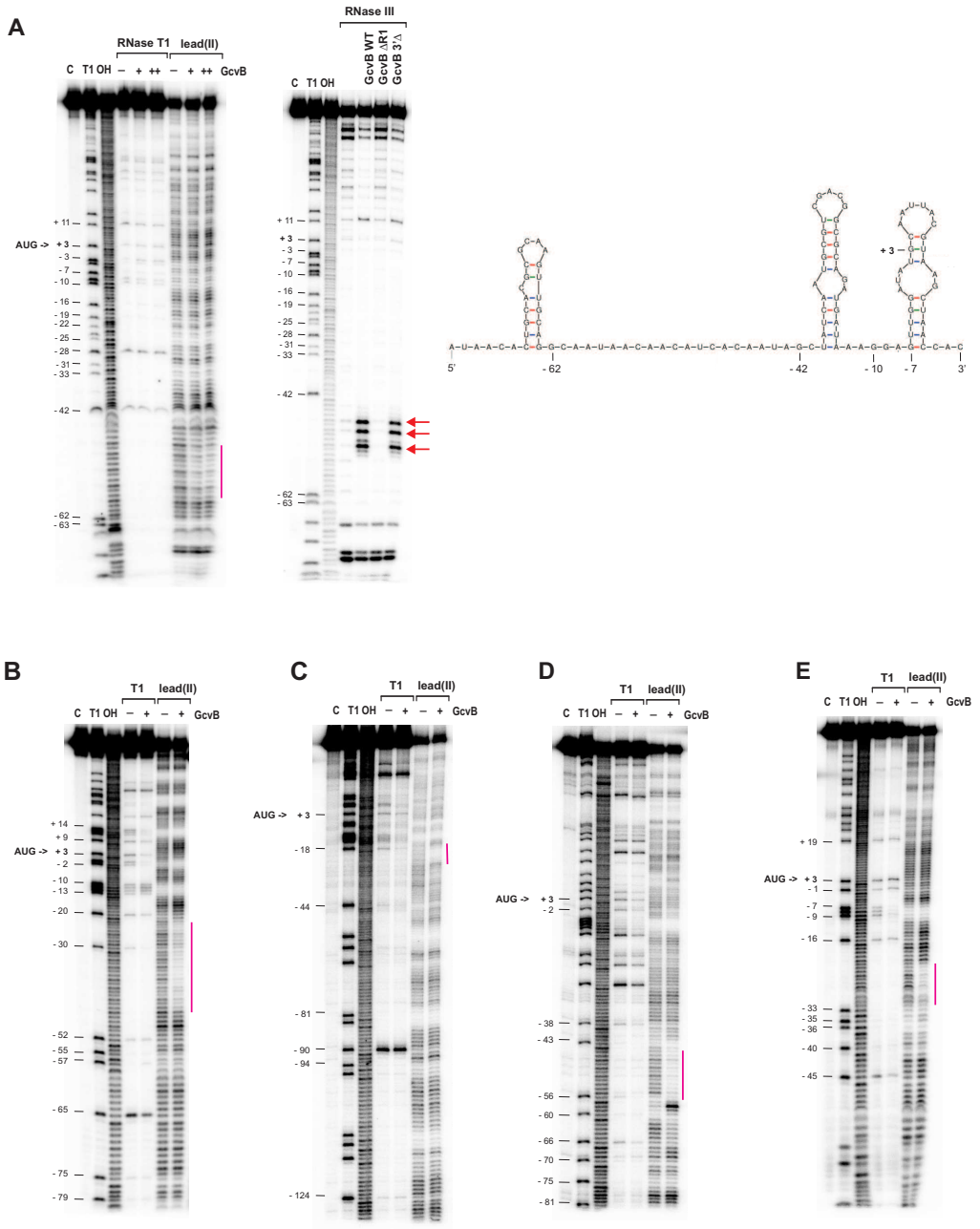


Figure 3.11: Footprinting of several periplasmic transport proteins mRNA leaders in complex with GcvB RNA. 5' end labelled target mRNA leader, (**A**, left panel) *gltI*, **(B)** *livJ*, **(C)** *livK*, **(D)** *argT*, and **(E)** STM4351, were subjected to RNase T1 and lead(II) cleavage, in the absence or presence (indicated by +) of cold GcvB RNA. Protected regions by GcvB RNA are indicated by pink bars and are located upstream of the start codon for all targets. Start codons are indicated by black arrows. Lane C: Untreated leader RNA. Lane T1: RNase T1 ladder of leader RNA under denaturing conditions. Positions of cleaved G residues in the target sequences are given as distance to the mRNA start codon. Position "+ 3" which is indicated in bold corresponds to the G residue of the start codon AUG. Lane OH: Alkaline ladder. GcvB binding site on the *gltI* leader RNA (**A** left) is supported by strong RNase III cleavages in this region in the presence of GcvB WT or 3'Δ mutant RNA (**A**, middle panel). In contrast, no cleavage is observed in the presence of mutant $GcvB_{\Delta R1}$ RNA, which lacks the G/U-rich region R1. The predicted secondary structure of the 5' *gltI* mRNA (**A**, right panel) was computed using the `mfold` algorithm (Zuker, 2003).

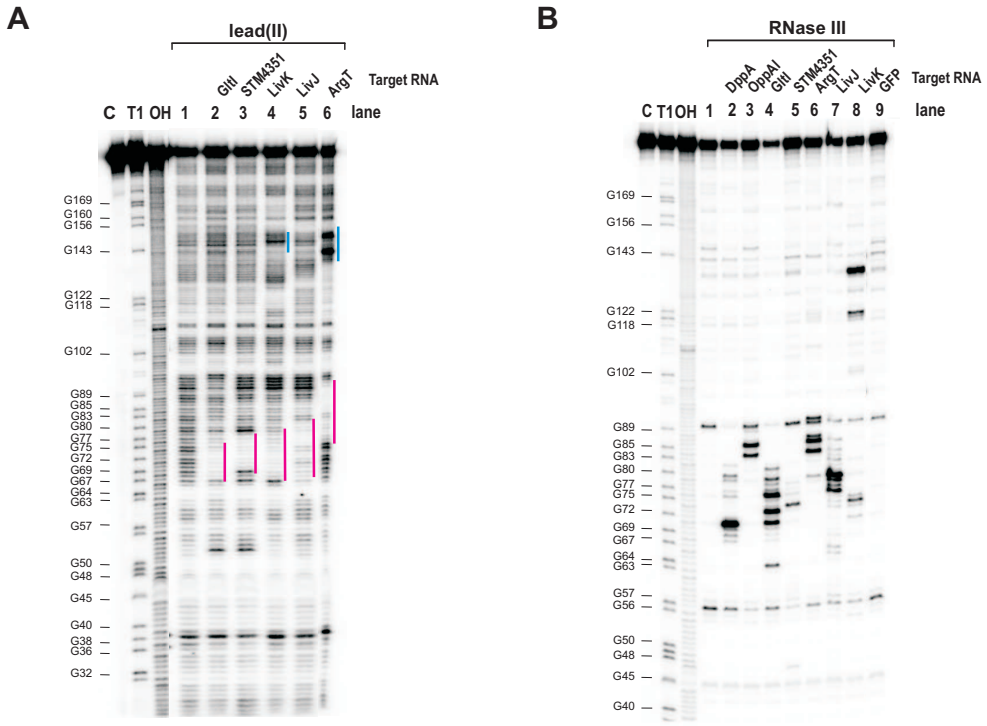


Figure 3.12: GcvB RNA binds to additional periplasmic transporter mRNAs via the GU-rich region R1. (A) 5' end-labelled GcvB RNA was subjected to lead (II) cleavage for 2 min at 37°C in the presence of different concentrations of periplasmic target mRNA leaders. Target mRNAs were added to the following final concentrations: GltI: 2 μM, STM4351: 500 nM, LivK: 500 nM, LivJ: 500 nM, and ArgT: 1 μM. These concentrations are based on preliminary gel shift assays with labelled GcvB RNA in combination with each target mRNA. For all targets a footprint in GcvB (indicated by a pink bar) is observed that is located in the GU-rich region R1 (position 66 - 89). Upon addition of LivK and ArgT leader also structural changes in the 3' end of GcvB (blue bars) could be observed. **(B)** RNase III cleavage after 7 min incubation at 37°C of labelled GcvB RNA in combination with different target mRNA leaders. For target mRNAs that are also present in **(A)** the same final concentrations were used; for DppA and OppA leader RNA 500 nM, and for GFP control RNA 1 μM were added. Addition of target RNA but not control GFP RNA leads to specific cleavage of GcvB RNA in the GU-rich region R1 and supports this region to be the target interaction site. Only addition of LivK leads to an additional cleavage site in the 3' part of GcvB RNA.

To examine regulation of the new targets *in vivo*, translational *gfp* fusions to all of these genes were constructed. The *gltI* and *livJ* fusions showed fluorescence on agar plates (Fig. 3.13A). Both fusions were strongly repressed by *gcvB* alleles with an intact R1 sequence (*gcvB*, $gcvB_{\Delta R2}$ and $gcvB_{3'\Delta T}$), whereas the $gcvB_{5'\Delta}$ and $gcvB_{\Delta R1}$ alleles lacking R1 failed to repress the fusions (Fig. 3.13A). Western blot analysis of all five target fusions as well as the *dppA::gfp* and *oppA::gfp* fusions further confirmed that regulation required an intact R1 sequence (Fig. 3.13B).

In addition, regulation of these targets by chromosomal GcvB was examined by quantitative real-time PCR analysis of target mRNA levels. A *Salmonella gcvB* deletion strain showed $\approx$8-fold higher *oppA* mRNA levels and $\approx$12-fold higher *dppA* mRNA levels, respectively. Also for the new targets, *gltI*, *livJ*, and STM4351, up to $\approx$6-fold higher mRNA levels were observed, indicating that

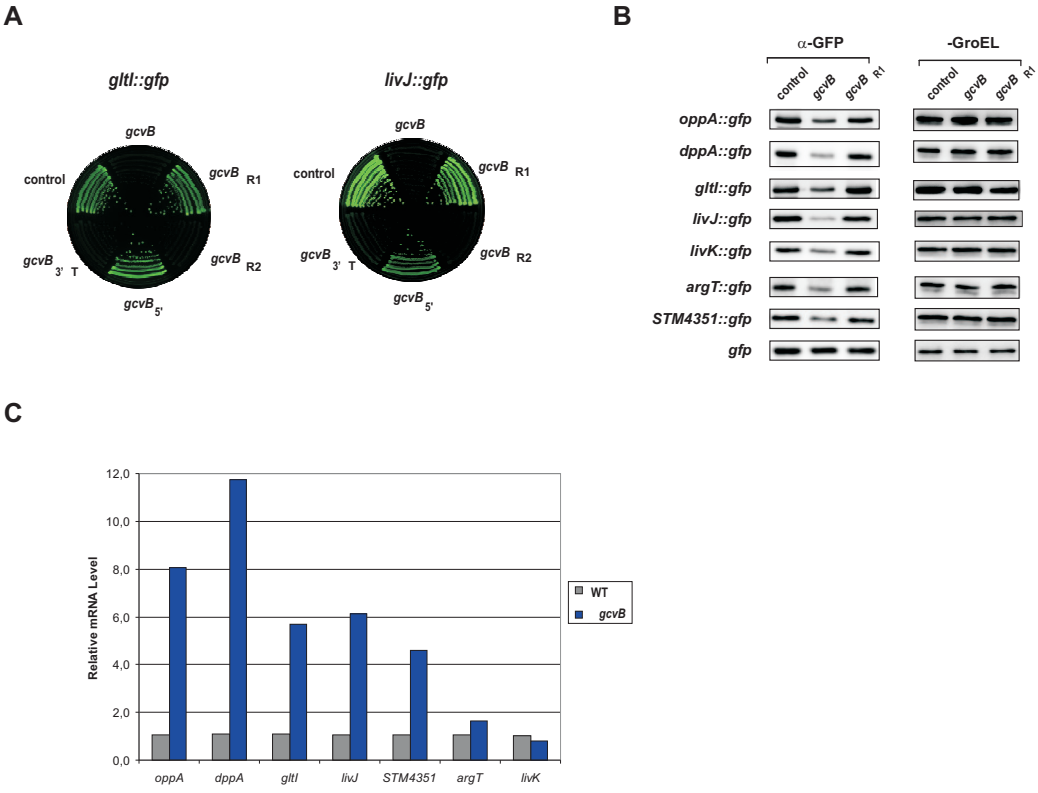


Figure 3.13: GcvB targets additional periplasmic transporter mRNAs. (A) Agar plate-based assay of colony fluorescence of *E. coli* strains carrying either *gltI::gfp* (*left*; plasmid pJL24-1) or *livJ::gfp* (*right*; plasmid pJL20-1). Each fusion plasmid was combined with control plasmid, pTP11, or plasmids expressing *Salmonella* GcvB and mutant RNAs as in Figure 3.4B. **(B)** Western blots of target::GFP fusion proteins prepared from *E. coli* Δ*gcvB recA*$^-$ carrying the indicated fusion plasmids in combination with control plasmid pTP11, or plasmids expressing wild-type *gcvB* or the $gcvB_{\Delta R1}$ allele. GroEL was probed as loading control. Fold changes of GFP fusion protein levels (upon normalization to GroEL levels) by *gcvB* or $gcvB_{\Delta R1}$ co-expression relative to the control plasmid were: OppA::GFP, -2.8/-1.2; DppA::GFP, -3.6/-1.3; GltI::GFP, -1.8/+2.2; LivJ::GFP, -5.3/-1.4; LivK::GFP, -2.1/-1.0; ArgT::GFP, -1.7/-1.1; STM4351::GFP, -1.8/-1.3; GFP alone, -1.1/-1.2. **(C)** Quantification of relative target mRNA levels using quantitative RT-PCR. Real-time PCR analysis of target mRNA levels in wild-type and Δ*gcvB Salmonella*. Samples were prepared from exponential phase bacteria (OD_{600} 0.4). Wild-type mRNA levels are set to 1 (grey bars), and blue bars show fold-upregulation in the absence of *gcvB*.

translational inhibition of these targets is coupled to mRNA degradation. Only for *argT* and *livK* similar mRNA levels were observed between *Salmonella* wild-type and the *gcvB* deletion strain, arguing for inhibition of translation as the primary mechanism of GcvB regulation for these targets. Overall, these results increased the number of GcvB targets to seven mRNAs.

3.1.5. GcvB inhibits translation initiation *in vitro*

GcvB binds near the Shine-Dalgarno sequence of *dppA* and *oppA* (Fig. 3.6D), and was thus predicted to prevent ribosome binding to these mRNAs. To test this, 30S ribosome toeprinting assays

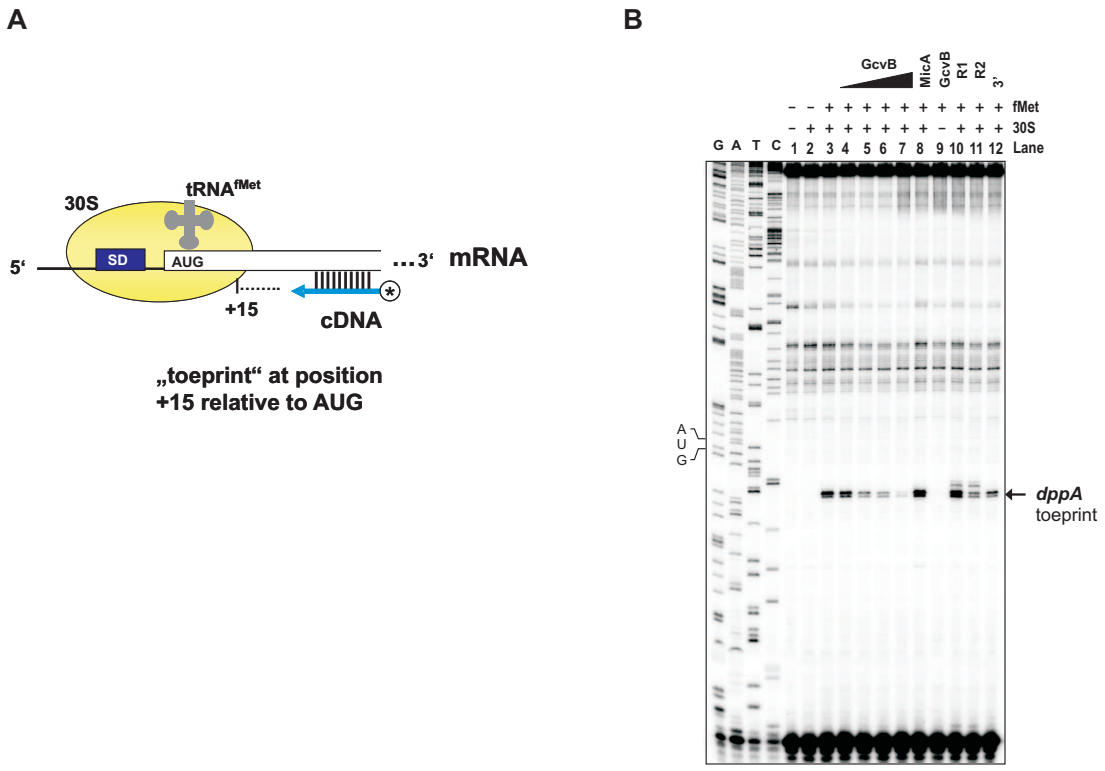


Figure 3.14: GcvB inhibits 30S binding *in vitro*. (A) Toeprinting assays: formation of ternary translation-initiation complexes of 30S subunit and fMet initiator tRNA on an *in vitro* transcribed mRNA are monitored in a reverse transcription reaction with a labelled antisense oligo. Only if both, 30S and $tRNA^{fMet}$, are bound to the SD and AUG start codon of the mRNA, reverse transcription stops at position '+15' according to the start codon and leads to the specific "toeprint" signal. **(B)** Ribosome toeprinting of *dppA* leader RNA (20 nM) as described in section 8.3.9 in Material and Methods. "+/-" indicate the presence or absence of 30S subunit (200 nM) and fMet initiator tRNA (1 μM). The *dppA* AUG start codon position is shown. The arrow indicates the 30S toeprint. Increasing concentrations of GcvB RNA (lanes 4-7: 20, 60, 100 and 200 nM) in the reactions inhibit 30S binding whereas the unspecific control RNA, MicA (lane 8, 100 nM), or $GcvB_{\Delta R1}$ mutant RNA (lane 10, 100 nM) do not impair binding. Mutant RNAs $GcvB_{\Delta R2}$ and $GcvB_{3'\Delta}$ (lanes 11 and 12) were added at a final concentration of 100 nM.

(Hartz *et al.*, 1988) were performed (Fig. 3.14A). A *dppA* mRNA fragment was annealed to an end-labelled primer complementary to the *dppA* coding region (+58 to +73), and incubated with 30S ribosomal subunits in the presence or absence of uncharged $tRNA^{fMet}$, followed by cDNA synthesis. Analysis of the extension products (Fig. 3.14B) revealed one ribosome-induced, $tRNA^{fMet}$-dependent termination site at the characteristic +15/+16 positions (start codon A is +1). This "toeprint" signal was decreased when wild-type GcvB RNA was added prior to incubation with 30S/fMet (lanes 4-7), suggesting inhibition of 30S binding. Inhibition was also observed with the $GcvB_{\Delta R2}$ and $GcvB_{3'\Delta}$ mutant RNAs, which have an intact R1 sequence (lanes 11, 12). In contrast, the $GcvB_{\Delta R1}$ mutant RNA (lane 10) did not inhibit ternary complex formation.

The same assay performed on *oppA* mRNA confirmed that GcvB inhibited ternary complex formation, and that it required an intact R1 sequence (data not shown). Since GcvB did not inhibit 30S

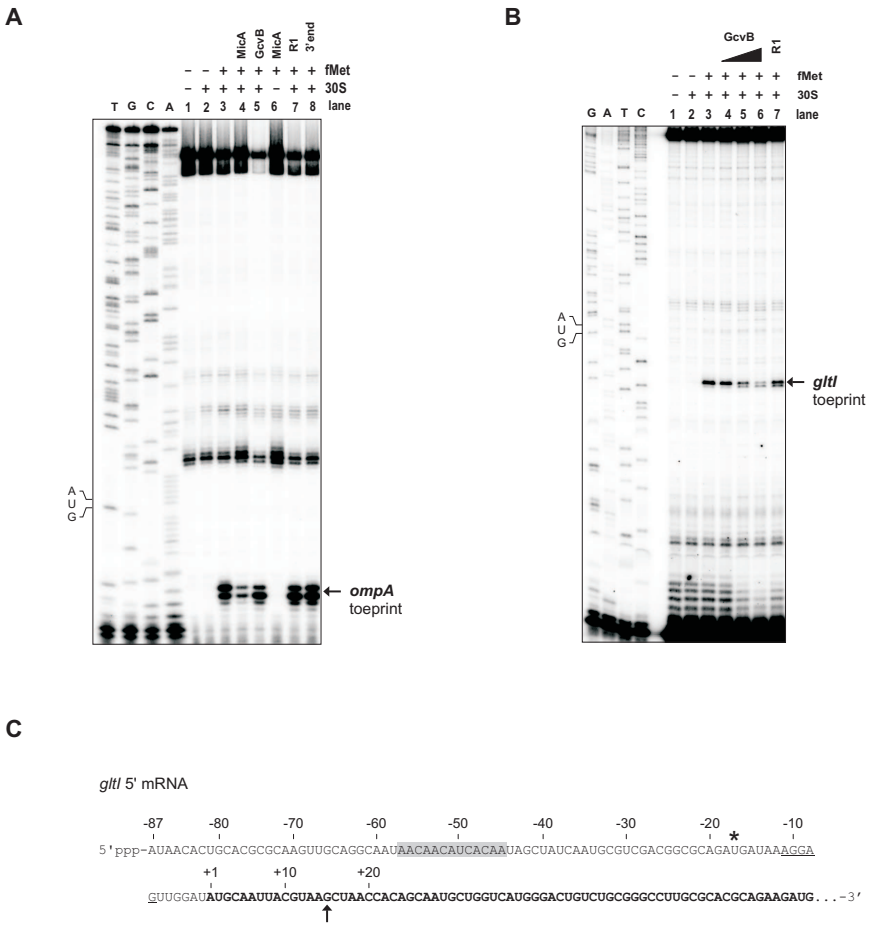


Figure 3.15: GcvB inhibits 30S binding to *gltI* mRNA from a distance. (A) Ribosome toeprinting of *ompA* leader RNA (200 nM) as described in section 8.3.9 in Material and Methods. "+/-" indicate the presence or absence of 30S subunit (200 nM) and fMet initiator tRNA (1 μM). The *ompA* AUG start codon position is shown. The arrow indicates the 30S toeprint. Only MicA (1μM) inhibits 30S binding to *ompA* mRNA as previously shown in Udekwu *et al.* (2005), whereas neither GcvB wild-type (1 μM) nor any of the GcvB mutants RNAs (1 μM) are able to interfere with translation initiation. **(B)** Toeprint on *gltI* 5' RNA (20 nM) using 20 nM 30S subunit and 100 nM $tRNA^{fMet}$. Increasing GcvB concentrations (lanes 4-6: 20, 100 and 200 nM, respectively) inhibit 30S binding. $GcvB_{\Delta R1}$ mutant RNA (100 nM, lane 7) does not inhibit the toeprint. **(C)** mRNA sequence of *Salmonella gltI* (5' end). The *gltI* coding sequence is set in bold-face. The asterisk denotes the wrongly annotated *gltI* start codon in the *Salmonella* LT2 genome sequence. The *gltI* start codon shown here was confirmed by a specific toeprint signal at position +15 which is indicated by an arrow (cf. **B**). The SD sequence is underlined, and the C/A-rich GcvB target site is shadowed.

binding to the unrelated *ompA* mRNA (see Figure 3.15A), it probably acts as a sequence-specific translation initiation inhibitor by masking the *dppA* and *oppA* RBS.

3.1.6. GcvB inhibits *gltI* translation by binding far upstream of the start codon

GcvB binds to its other five target mRNAs further upstream of SD and start codon as compared to *dppA* and *oppA* (Fig. 3.10). Notably, GcvB binds to region -57 to -45 of *gltI*, and -57 to -42 of

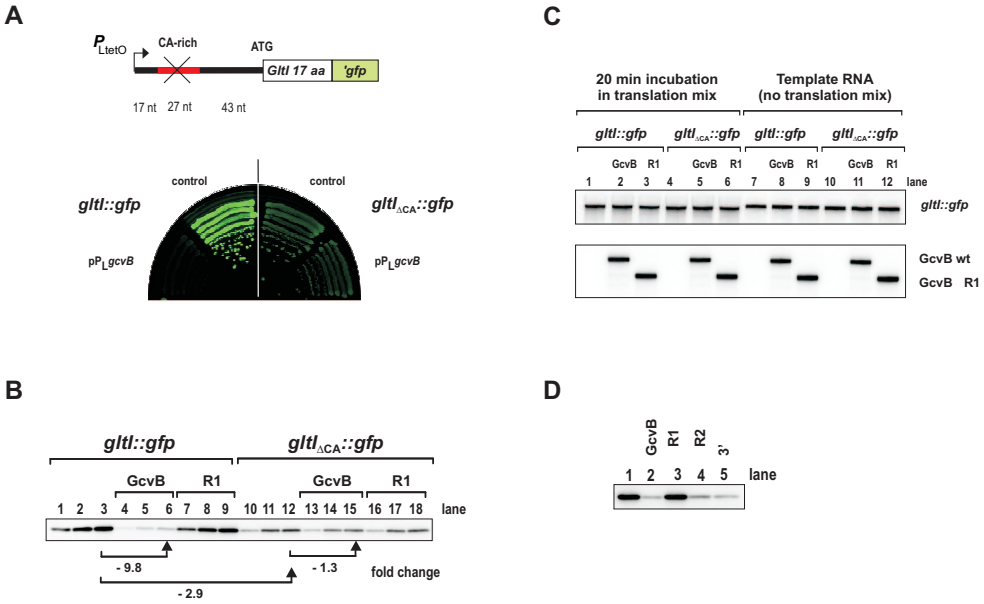


Figure 3.16: The C/A-rich upstream element functions both as GcvB target site and as translational enhancer element. **(A)** Agar plate-based assay of colony fluorescence of *E. coli* strains carrying either the parental *gltI::gfp* (*left*; plasmid pJL24-1) or the mutant *gltI*$_{\Delta CA}$*::gfp* (*right*; plasmid pJL45-3) fusion. In the latter plasmid, the C/A-rich GcvB target site (-71 to -44; cf. Fig. 3.15C) is deleted. The fusion plasmids were combined with control vector pTP-11, or GcvB expression plasmid pP_L*gcvB*. **(B)** *In vitro* synthesized, full-length mRNAs (40 nM) of the *gltI::gfp* and *gltI*$_{\Delta CA}$*::gfp* fusions were *in vitro* translated with reconstituted 70S ribosomes in the presence of 40 nM Hfq as described in section 8.3.10 in Material and Methods. Synthesis of GltI::GFP fusion protein levels were determined after incubation for 15 min (lanes 1, 4, 7, 10, 13, 16), 30 min (lanes 2, 5, 8, 11, 14, 17), and 45 min (lanes 3, 6, 9, 12, 15, 18) by Western blot analysis. The reactions in lanes 4-9 and 13-18 contained a 10-fold excess of GcvB or GcvB$_{\Delta R1}$ mutant RNAs as indicated above the lanes. Fold-regulations upon deletion of the C/A-rich target site or by addition of GcvB according to quantification of GltI::GFP fusion protein levels are given below the blot. The results of a representative experiment (out of three) are shown. **(C)** Confirmation of RNA stability in *in vitro*-translation reactions: 0.4 pmol of *in vitro* synthesized, full-length mRNAs of the *gltI::gfp* and *gltI*$_{\Delta CA}$*::gfp* fusions were translated with reconstituted 70S ribosomes in the presence of 0.4 pmol Hfq as described in Section 8.3.10 under Material and Methods. RNA samples of all reactions were prepared after translation for 20 min (lanes 1 to 6), and analysed by Northern hybridization along with untreated control RNAs (lanes 7 to 12). Probing for fusion mRNA or GcvB RNA with labelled oligos JVO-0155 and JVO-0749, respectively, shows that RNAs are stable in this assay. **(D)** Full-length *gltI::gfp* mRNA was *in vitro* translated for 30 min as above and synthesis of GltI::GFP fusion protein monitored on Western blot. In addition, mutant RNAs GcvB$_{\Delta R1}$ and GcvB$_{3'\Delta}$ (rightmost two lanes; 10-fold excess) were included in the assay.

argT. These interaction sites lie outside the -35 to +19 mRNA region covered by 30S ribosomes (Hüttenhofer & Noller, 1994). Toeprinting of 30S on *gltI* mRNA (Fig. 3.15B) gave a single toeprint at +15 relative to the *gltI* AUG in the presence of 30S/fMet (lane 3), confirming the *gltI* start codon shown in Figure 3.15C (note that the *gltI* coding region is misannotated in the *Salmonella* genome, and likely in a number of other bacteria). Strikingly, GcvB dramatically reduced 30S binding on the *gltI* mRNA (lanes 4 to 6), dependent on an intact R1 sequence (lane 7). Thus, GcvB can inhibit mRNA translation by binding upstream of a 30S binding site.

3.1.7. The C/A-rich target site enhances *gltI* translation

To further evaluate GcvB regulation at upstream mRNA sites, the C/A-rich target region (-71 to -44) was deleted in *gltI::gfp*, yielding fusion *gltI*$_{\Delta CA}$*::gfp*. This deletion abolished fusion repression in the presence of pP_L *gcvB* (Fig. 3.16A), further demonstrating that GcvB pairs to this region. Intriguingly, loss of this region also resulted in lower GFP fluorescence as compared to the parental *gltI::gfp* fusion. Since C/A-rich stretches were reported to enhance mRNA translation (Martin-Farmer & Janssen, 1999), it was determined whether the GcvB target site acts as an enhancer of *gltI* translation.

In vitro synthesized, full-length *gltI::gfp* and *gltI*$_{\Delta CA}$*::gfp* fusion mRNAs were translated with reconstituted 70S ribosomes and GltI::GFP fusion protein synthesis was monitored over time. Figure 3.16B shows a linear increase in GltI::GFP synthesis from both mRNA templates within a 45 min assay. However, the *gltI::gfp* mRNA gave ≈3-fold higher translation rates compared to that of *gltI*$_{\Delta CA}$*::gfp*. Addition of GcvB (tenfold excess) to the reactions strongly inhibited protein synthesis from *gltI::gfp* but not *gltI*$_{\Delta CA}$*::gfp* mRNA. None of these variations are due to RNA degradation since the *gltI::gfp* and GcvB RNAs were stable during the assay (Fig. 3.16C). In addition, the inhibitory activity of various GcvB mutants on *gltI::gfp in vitro* (Fig. 3.16D) perfectly correlated with their ability to inhibit expression of the fusion mRNA *in vivo* (Fig. 3.13A).

3.1.8. Creation of an upstream C/A-rich target site permits translational control of an unrelated mRNA

Next, the C/A-rich GcvB binding site of *gltI* was inserted at position -42 of an *E. coli ompR::gfp* fusion, yielding *ompR*$_{CA}$*::gfp* (Fig. 3.17A). The *Salmonella gltI* and *E. coli ompR* genes are unrelated in function, and their 5' mRNA regions have little sequence identity (Fig. 3.17B). Insertion of the C/A-rich element resulted in a ≈2-fold increase in OmpR::GFP synthesis in the *in vitro* translation assay (Fig. 3.17C; compare lanes 3 and 9), confirming the stimulatory effect of this element on mRNA translation. When GcvB was present in the reaction, it strongly inhibited translation of the *ompR*$_{CA}$*::gfp* but not of the parental *ompR::gfp* fusion mRNA (Fig. 3.17C). Intriguingly, GcvB inhibits *ompR*$_{CA}$*::gfp* translation ≈ 10-fold, *i. e.* as strongly as observed with *gltI::gfp*. In other words, GcvB effectively represses a structurally unrelated mRNA upon insertion of a C/A-rich target site at an upstream position.

3.2. Discussion

ABC transporters constitute a major class of amino acid uptake systems and commonly have at least one periplasmic solute binding protein to take up substrates upon their diffusion through outer membrane porins (Chen *et al.*, 2004; Hosie & Poole, 2001). Given the key roles of amino acids in nitrogen and carbon metabolism, regulation of transporter expression and activity is complex. Intriguingly, genes involved in amino acid uptake and metabolism also are the most strongly misregulated class in *hfq* mutants *E. coli* and *Salmonella* (Guisbert *et al.*, 2007; Sittka *et al.*, 2008), suggesting that Hfq-dependent sRNAs are involved in the post-transcriptional control of these pathways. Here, one such sRNA, GcvB, has been established as a direct regulator of many ABC trans-

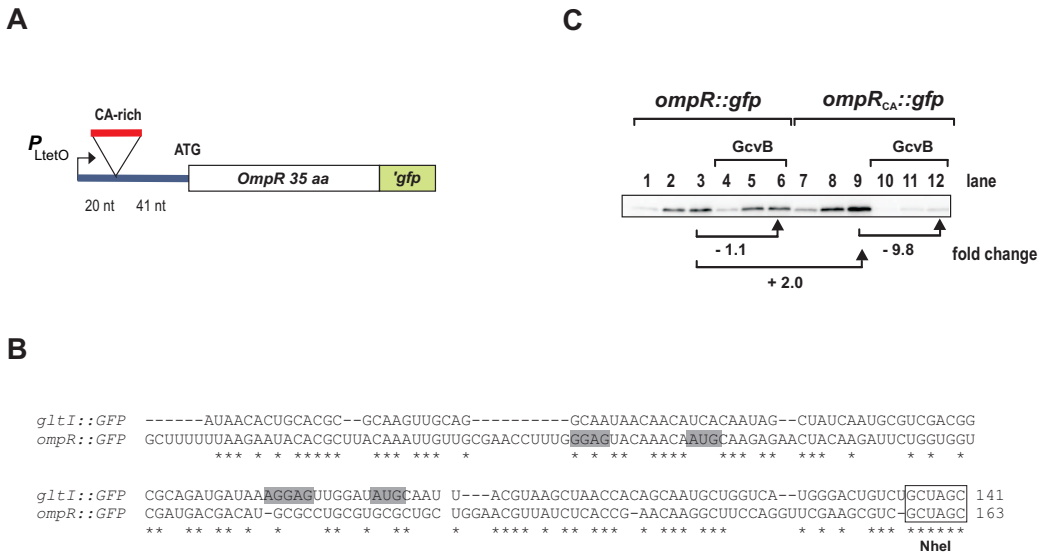


Figure 3.17: Transplantation of the C/A element to an unrelated mRNA enhances translation and permits regulation by GcvB. (A) Insertion of the 27 nt C/A-rich element of *gltI* at position -42 relative to the start codon of an *E. coli ompR::gfp* fusion yields mutant *ompR*$_{CA}$*::gfp* (plasmids pJU-063 and pJL50-11, respectively). **(B)** *In vitro* translation assay (reconstituted 70S ribosomes) as in Figure 3.16B but with *in vitro*-synthesized *ompR::gfp* and *ompR*$_{CA}$*::gfp* fusion mRNAs (250 nM) and in the presence of 250 nM Hfq. Synthesis of OmpR::GFP fusion protein was determined on Western blots following incubation for 15 min (lanes 1, 4, 7, 10), 30 min (lanes 2, 5, 8, 11), and 45 min (lanes 3, 6, 9, 12). A tenfold excess of GcvB RNA over mRNA was added to the reactions where indicated. The results shown are representative of the experiment performed in triplicate. **(C)** `Clustalw`[1] alignment of *E. coli ompR* and *Salmonella gltI* mRNAs (5' parts, which are contained in the *gfp* fusions) shows that these RNAs are unrelated in sequence. SD sequences and start codons are shadowed.

porter mRNAs in *Salmonella*. The conservation of *gcvB* genes combined with the previous data from the Stauffer lab (McArthur *et al.*, 2006; Urbanowski *et al.*, 2000) as well as successful prediction of GcvB-target interactions in related organisms (see Figures 10.2 to 10.6 in the Appendix and Tjaden *et al.*, 2006) strongly argue that the GcvB function in amino acid uptake is conserved in many other bacteria. Interestingly, some enterobacteria encode an additional sRNA, RydC, which regulates the expression of a putative ABC transport system, *yejABEF* (Antal *et al.*, 2005).

GcvB directly interacts with seven *Salmonella* mRNAs that belong to various transcriptional regulons but collectively encode periplasmic substrate binding proteins of ABC transporters (`TCDB`[2]; `KEGG`[3]). These proteins are known or predicted to bind di- and oligopeptides (DppA, OppA), polar (GltI, STM4351, and ArgT) and branched amino acids (LivK, LivJ). Since GcvB is specifically expressed in rich medium in fast-growing cells (Fig. 3.1B), its function appears to lie in the general repression of amino acid uptake when nutrients are plentiful. Consistent with this prediction, a *Salmonella gcvB* deletion strain exhibits elevated protein and mRNA levels of almost all established GcvB targets (Figs. 3.3B, 3.9, and 3.13C). The increased steady-state levels of nearly all

[2] `http://www.tcdb.org/`

[3] `www.kegg.org`

target mRNAs in Δ*gcvB* (Fig. 3.13C) argue that by repressing mRNA translation, GcvB also promotes the decay of its targets. Western blot quantification of OppA protein levels revealed a 6-fold increase in Δ*gcvB* as compared to wild-type (Fig. 3.3B); the corresponding RT-PCR experiment (Fig. 3.13C) suggests a similar increase in *oppA* mRNA levels. Comparison of the increases in target protein levels (spot intensity on Coomassie-stained 2D gels; Fig. 3.9) and mRNA level (Fig. 3.13C) upon *gcvB* deletion suggest a similar correlation in mRNA/protein level changes for *dppA*, *gltI*, *livJ*, STM4351, and *argT* (LivK was not detected on the 2D gels). Whether RNase III, here used as a tool for structure probing *in vitro*, is the primary nucleolytic activity to degrade the GcvB target mRNA *in vivo* as previously shown for the IstR-1 sRNA (Vogel *et al.*, 2004) and RNAIII (Huntzinger *et al.*, 2005) remains to be determined. At least *dppA*, *livK*, *gltI*, and *oppA* each constitute the first gene of a polycistronic transporter operon. Considering that the downstream cistrons of these polycistronic mRNAs are likely to be translationally coupled to the direct GcvB targets, the sRNA may in fact repress >20 genes (see Section 4.1.2 in Chapter 4). However, since GcvB is intrinsically unstable (half-life 2-3 min; Vogel *et al.*, 2003 and Figure 3.3), ABC transporter repression will be quickly alleviated upon decreased availability of amino acids.

The sheer number of sRNA genes discovered by systematic searches of bacterial genomes over the last years led to an increasing recognition of the potential impact of these riboregulators on bacterial physiology (see Section 2.3). It is well-established that sRNAs that act to modulate protein activity can control the expression of many genes, *e. g.*, by binding to RNA polymerase (Wassarman & Storz, 2000) or to CsrA-like proteins (Babitzke & Romeo, 2007). Small RNAs can also affect the expression of larger sets of genes by targeting the mRNAs of global transcriptional regulators, *e. g.*, RpoS or FhlA (Argaman & Altuvia, 2000; Lease *et al.*, 1998; Majdalani *et al.*, 1998). In contrast, the understanding of how sRNAs could directly control multiple mRNAs by antisense mechanism has been limited by the low number of validated sRNA-target interactions and hence the difficulty to reliably predict new targets. This notwithstanding, the results of several recent studies suggest that multiple targeting may be more common than previously thought, and common denominators are beginning to emerge. Tjaden *et al.* (2006) suggested that functional relationship of the proteins encoded by target candidates, as it is also seen among the GcvB targets, may add confidence to predictions. Similarly, the OmrA/B and RybB sRNAs directly target multiple mRNAs that collectively encode for outer membrane proteins (Bouvier *et al.*, 2008; Guillier & Gottesman, 2006, 2008; Johansen *et al.*, 2006; Papenfort *et al.*, 2006). By the same token, the iron stress-responding *E. coli* RyhB sRNA was shown to regulate multiple mRNAs that encode proteins involved in iron metabolism (Massé & Gottesman, 2002), and many of the mRNAs demonstrated to be direct targets of *Staphylococcus aureus* RNAIII encode bacterial virulence factors (Boisset *et al.*, 2007).

The strong C/A bias of GcvB binding sites may point to yet another feature of multiple target regulation. CA multimers placed downstream of mRNA start codons were reported to stimulate translation *in vivo*, and to increase ribosome binding affinity to mRNAs *in vitro* (Martin-Farmer & Janssen, 1999). Whilst the effect of CA-multimers far upstream of a start codon was not addressed by these authors, here it turned out that the C/A-rich element has a stimulatory effect on the *gltI* and *ompR* fusion mRNAs (Figs. 3.16 and 3.17). Thus, multiple mRNA targeting by GcvB may have evolved through hijacking a translational enhancer element shared by numerous mRNAs that encode periplasmic transporters. It is thus tempting to speculate that other evolutionary constrained mRNA regions, *e. g.*, those encoding signal peptides, may also constitute binding sites for sRNAs

with multiple targets. However, it has to be emphasized that the effects of GcvB as a translational repressor are clearly much greater than the effects of removing the C/A-rich target site (Fig. 3.16), thus GcvB does not primarily act by simply blocking any enhancer effect.

In eukaryotes, the recognition that a 5' terminal seed sequence of the ≈ 22 nt microRNAs provides the base-pairing specificity to target mRNAs greatly advanced target predictions. Also some of the much longer bacterial sRNAs, *e. g.*, OmrA/B and RybB, have recently been shown to have highly conserved 5'-terminal binding sites for recognition of their target mRNAs (Bouvier *et al.*, 2008; Guillier & Gottesman, 2008). However, the conserved target interaction sites of sRNAs are not necessarily located at the 5' end. An alignment of GcvB RNAs (Fig. 3.2C) revealed two strongly conserved, internal regions, R1 and R2. Whilst R2 may constitute an interaction region for yet another target(s), R1 is a key determinant for base-pairing to seven target mRNAs. Although the structure probing of target mRNAs and GcvB indicates additional, weaker contacts, the C/A-rich motif in the targets and R1 in the sRNA clearly are the anchoring and essential motifs for interaction. Recently, a highly conserved, internal target site has also been observed, *e. g.*, for CyaR RNA (De Lay & Gottesman, 2009; Johansen *et al.*, 2008; Papenfort *et al.*, 2008). Accordingly, giving higher weight to deeply conserved regions in sRNAs is expected to improve the currently available target search algorithms. In support of this, several other regulatory sRNAs, *e. g.*, MicA, MicC, and SgrS, exhibit a higher degree of conservation in their target interaction regions (Chen *et al.*, 2004; Rasmussen *et al.*, 2005; Udekwu *et al.*, 2005; Vanderpool & Gottesman, 2004). Unlike these enterobacterial sRNAs, *S. aureus* RNAIII is not conserved in other species. However, it is still intriguing that a short (≈ 40 nt) region within the 514 nt RNAIII facilitates base-pairing to multiple mRNAs (Boisset *et al.*, 2007; Huntzinger *et al.*, 2005). Different from GcvB, however, RNAIII typically covers the SD and/or the start codon of a target mRNA (Boisset *et al.*, 2007; Huntzinger *et al.*, 2005).

Competition with ribosome binding explains the inhibitory activity of sRNAs that bind mRNAs within RBS regions (Argaman & Altuvia, 2000; Chen *et al.*, 2004; Huntzinger *et al.*, 2005; Udekwu *et al.*, 2005). Here it is shown that the same applies to GcvB inhibition of *dppA* and *oppA*, whose GcvB binding sites are very close to the SD (Fig. 3.6D). However, GcvB binds many of its targets further upstream of the two sequence elements, SD and start codon, that are the key determinants of 30S binding (Fig. 3.10B). *gltI* is the mRNA with the most upstream binding site; probing of GcvB/*gltI* RNA complexes as well as transfer of the GcvB site from *gltI* to the unrelated *ompR* mRNA showed that GcvB effectively represses translation by forming a duplex more than 42 nt upstream of the AUG.

How is translational repression brought about at such upstream sites? Footprinting experiments revealed the maximal ribosome binding region on mRNAs to range from -39 to +19 relative to the AUG (Hüttenhofer & Noller, 1994). Since the GcvB binding site on *gltI* lies outside this window, a simple interference model in which GcvB occludes mRNA residues required for base-pairing with 16S ribosomal RNA is unlikely. In eukaryotes, ribosomes generally enter mRNAs at their 5' end to subsequently scan for downstream AUG triplets. A requirement for scanning is unknown in prokaryotes, which argues against the possibility that GcvB could act as a roadblock for ribosomes scanning from the 5' end of *gltI* mRNA.

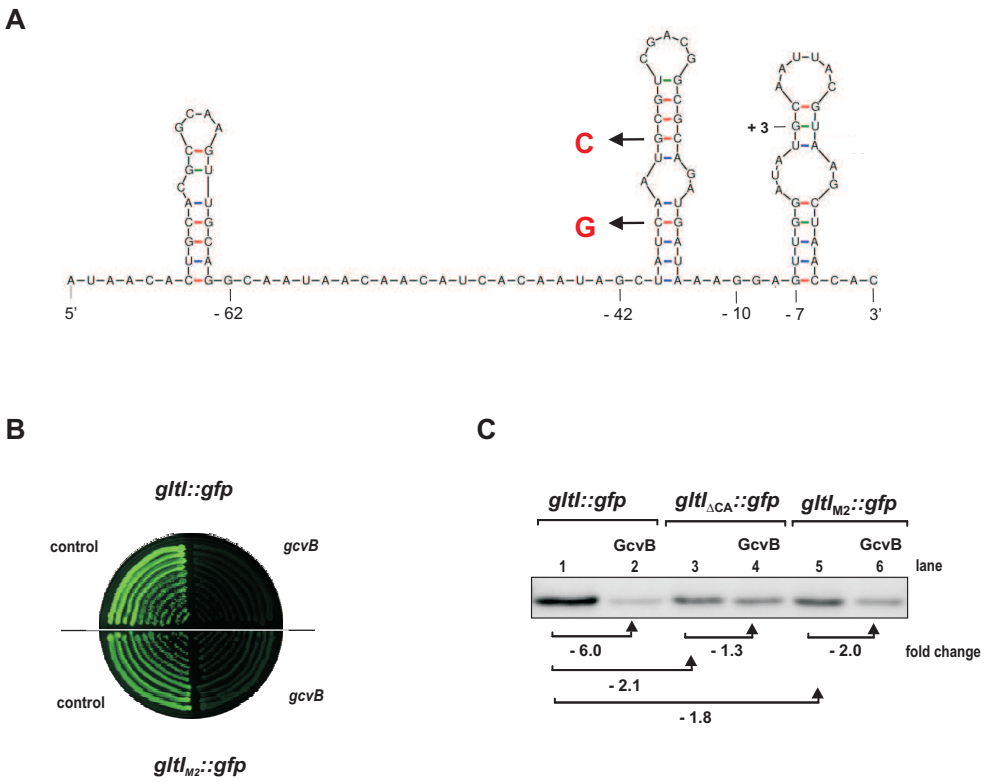


Figure 3.18: Disruption of the hairpin motif upstream of the *gltI* SD does not impair regulation of *gltI::gfp* by GcvB. (A) The *gltI* leader can form a hairpin structure (see also Figure 3.11A) which could bring the C/A-element closer to the SD sequence. This stem-loop was disrupted by mutation of two indicated residues (M2 mutation: C_{-37}→G and G_{-33}→C). Successful disruption of this hairpin by the M2 mutation was confirmed by RNA structure probing (data not shown). **(B)** Agar plate-based assay of colony fluorescence of *E. coli* strains carrying either the parental *gltI::gfp* (*top*; plasmid pJL24-1) or the mutated *gltI*$_{M2}$*::gfp* (*bottom*; plasmid pJL56-2) fusion. The fusion plasmids were combined with control vector pTP-11, or GcvB expression plasmid P_L *gcvB*. Disruption of the hairpin does not impair fusion mRNA repression by GcvB. **(C)** *In vitro* synthesized, full-length mRNAs (40 nM) of the *gltI::gfp*, *gltI*$_{\Delta CA}$*::gfp*, and *gltI*$_{M2}$*::gfp* fusions were *in vitro* translated with reconstituted 70S ribosomes in the presence of 40 nM Hfq as described in Section 8.3.10 under Material and Methods. Synthesis of GltI::GFP fusion protein levels were determined after incubation for 45 min by Western blot analysis. The reactions in lanes 2, 4, and 6 contained a 10-fold excess of GcvB RNAs as indicated above the lanes. Fold-regulations upon deletion of the C/A-rich target site, introduction of the M2 mutations or by addition of GcvB according to quantification of GltI::GFP fusion protein levels are given below the blot.

Two other bacterial sRNAs, IstR-1 and RyhB, were recently reported to repress translation by binding upstream of the target RBS (Darfeuille *et al.*, 2007; Vecerek *et al.*, 2007), yet the underlying mechanisms do not seem to apply for GcvB either. IstR-1 targets the *tisAB* mRNA ≈ 100 nt upstream of the *tisB* start codon. The *tisAB* mRNA is highly structured and the *tisB* SD entrapped in a stable hairpin. Thus, *tisB* translation requires ribosomes to bind to an upstream "standby" site, which will be masked upon IstR-1 binding (Darfeuille *et al.*, 2007). Unlike in *tisAB*, the *gltI* 5' UTR is not strongly structured (Fig. 3.11A). Moreover, since GcvB also inhibits the unrelated *ompR*$_{CA}$*::gfp* mRNA, structure cannot be a primary cause of translational repression. RyhB inhibits

fur translation by forming an imperfect duplex with the -96 to -48 region of *fur* mRNA. Translation of *fur* is coupled to that of an upstream located reading frame, the direct target of RyhB (Vecerek *et al.*, 2007). Here no evidence was found for an upstream reading frame in the *gltI* mRNA, nor for translation initiation upstream of the determined *gltI* start codon (Fig. 3.15B).

The *gltI* 5' UTR contains a putative hairpin motif (-40 to -7 region), which could bring the C/A-rich motif in closer proximity to the *gltI* SD to facilitate repression by GcvB (Fig. 3.18). The following observations do not support such a model. Firstly, the SD of *ompR::gfp* is not preceded by a hairpin that would bring the GcvB target site and SD in closer proximity; yet transfer of the GcvB site to *ompR::gfp* yields the same degree of mRNA repression as observed with the *gltI::gfp* construct. Secondly, the putative hairpin motif in the *gltI* 5' UTR was disrupted by two point mutations. These mutations did not impair GcvB's ability to repress the *gltI* fusion mRNA *in vivo* (Fig. 3.18, B). In contrast, in an *in vitro* translation assay, both mutant mRNAs, $gltI_{\Delta CA}$*::gfp* and $gltI_{M2}$*::gfp*, show $\approx$ 2-fold lower translation rates compared to that of *gltI::gfp* (Fig. 3.18, C). Furthermore, the ability of GcvB to repress translation of the $gltI_{M2}$*::gfp* fusion mRNA is $\approx$ 3-fold reduced compared to repression of *gltI::gfp* (Fig. 3.18, C). Thus, the hairpin structure could have an impact on GcvB regulation within the sequence and structure context of *gltI* mRNA. The observation, that the two mutations had no influence on regulation of the *gltI::gfp* fusion *in vivo* could be due to the overexpression of GcvB RNA from the plasmid. High GcvB levels could facilitate that GcvB can still bind the *gltI::gfp* fusion mRNA and lead to mRNA degradation by a double-strand specific RNase rather than inhibition of translation. Introduction of the two mutations into the chromosomal copy of *gltI* in the *Salmonella* wild-type strain and *gcvB* deletion strains and subsequent analysis of mRNA levels by quantitative RT-PCR in both strains, could help to clarify the actual *in vivo* situation.

Consequently, the current results suggest that either the first three GcvB stem-loops, or the G/U:C/A-rich helix of the GcvB-target mRNA complex, constitute an inhibitory signal for 30S entry on the *gltI* and *argT* mRNAs. However, the underlying mechanism may be a more general one since it was found that *Salmonella* RybB sRNA also represses several targets by binding upstream of the RBS (F. Mika and J. Vogel, unpublished). It might therefore be of interest to consider the possibility that target searches are not yet exhausted for those sRNAs that do not show obvious complementarity to RBS sequences.

In summary, *Salmonella* GcvB RNA was shown to target C/A-rich elements in the 5'UTRs of seven periplasmic ABC transporter mRNAs (see Fig. 3.19). A highly conserved G/U-rich region R1 within GcvB was identified which is strictly required for regulation and mediates direct binding to the C/A-rich target sites. The GcvB target sites are located inside and upstream of the ribosome binding site of the target mRNAs. However, in both cases, GcvB binding leads to inhibition of translation by interfering with ribosome binding. Furthermore, the C/A-rich elements were shown to act as translational enhancers in the 5' UTRs of the ABC transporter mRNAs. Thus GcvB RNA has probably captured an evolutionarily conserved element to regulate a class of functionally related genes.

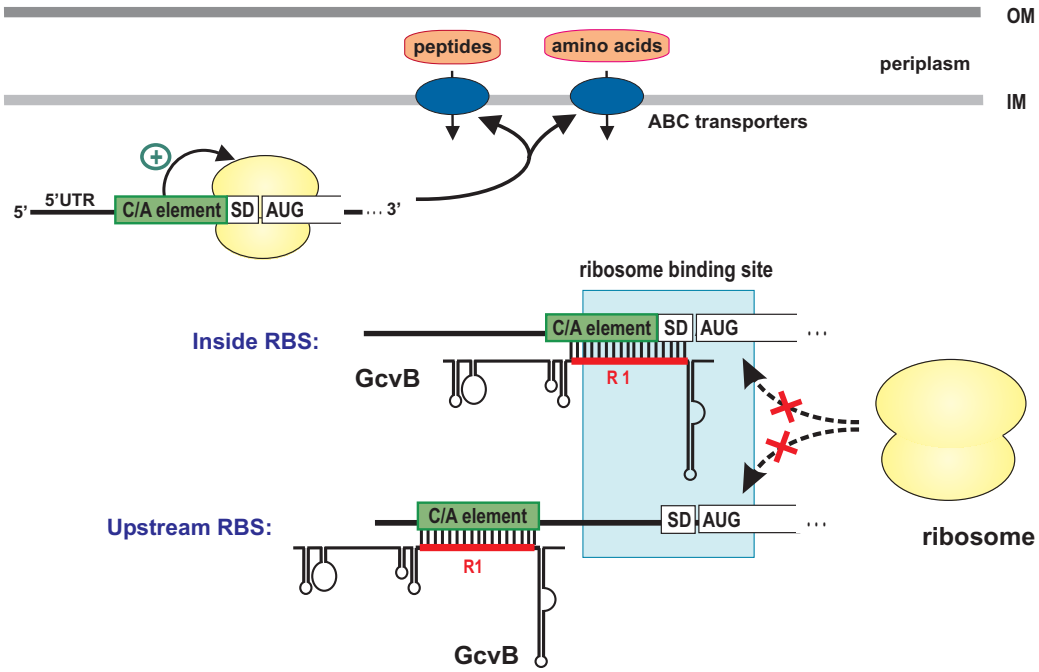


Figure 3.19: GcvB RNA targets multiple ABC transporter mRNAs inside and outside of ribosome binding sites. GcvB RNA binds with its highly conserved G/U-rich region R1 (red bar) to C/A-rich target sites (green boxes) inside and upstream of the ribosome binding sites (indicated by a blue box). In both cases, GcvB binding leads to inhibition of translation by interfering with ribosome binding. All seven periplasmic ABC transporter mRNAs which are repressed by GcvB carry a C/A-rich target site within their 5' UTR which in addition acts as a translational enhancer.

CHAPTER 4

GcvB RNA, A GLOBAL REGULATOR OF GENES INVOLVED IN AMINO ACID METABOLISM

In the previous Chapter, expression of seven ABC transporter mRNAs was shown to be repressed at the post-transcriptional level by the small noncoding RNA, GcvB, of *Salmonella*. GcvB directly binds with a highly conserved G/U-rich region to C/A-rich elements in the 5' UTRs of all these target mRNAs. This initial set of GcvB targets was based on proteomic analysis and bioinformatic predictions of interactions with 5' regions (-/+ 50 nt of start codon) of mRNAs. However, in both approaches mainly periplasmic proteins were investigated based on known targets from *E. coli*. The pleiotropic nature of *E. coli* and *Yersinia gcvB* mutants (McArthur *et al.*, 2006; Urbanowski *et al.*, 2000) and biocomputational predictions (Tjaden *et al.*, 2006) suggested that GcvB may have additional mRNA targets. Furthermore, there is growing evidence that sRNAs regulate multiple targets rather than individual mRNAs, thereby reprogramming gene expression at the post-transcriptional level (as outlined in Section 2.5).

Global approaches using co-immunoprecipitation coupled to microarray (Zhang *et al.*, 2003) or deep sequencing analysis (Sittka *et al.*, 2008; see Chapter 5) have identified numbers of Hfq-associated mRNAs that surpass those of known sRNAs, *e. g.*, $\geq$700 mRNAs versus $\approx$100 sRNAs in *Salmonella*. These Hfq-bound mRNAs represent potential sRNA targets, but it remains difficult to identify direct sRNA-mRNA partners. The few bioinformatics based approaches for bacterial sRNA target predictions (see Section 2.4.2) often predict many targets including a high number of false positives. Based on the low number of biochemically-verfied interactions, they often rely on the canonical model which considers sRNA pairing to the Shine-Dalgarno (SD) or AUG start codon sequence as a hallmark of productive target repression. However, several groups have recently reported regulation outside this narrow and low sequence complexity region (Bouvier *et al.*, 2008; Darfeuille *et al.*, 2007; Sharma *et al.*, 2007; Vecerek *et al.*, 2007); these studies have expanded the sequence space for productive targeting on the mRNA side. Furthermore, sRNA-mediated pull-down of interacting mRNAs (Douchin *et al.*, 2006) or microarray-based experimental approaches using sRNA pulse-expression (see Section 2.4.1.2) have predicted with high confidence diverse sRNAs to directly regulate more than one mRNA. The latter strategy is likely to avoid the pleiotropic effects that can be expected to result from constitutive sRNA expression and has been successfully applied to the identification of targets of several sRNAs, such as RyhB and RybB, as well as OmrA and OmrB (see Section 2.4.1.2).

This Chapter describes an extended search for additional GcvB targets by global analysis of mRNA changes on microarrays after GcvB pulse-overexpression. The analysis included GcvB mutants, lacking the conserved regions R1 or R2, to identify potential targets that are dependent on these regions. Additional consensus R1-dependent targets were identified by bioinformatics-based predictions of C/A-rich GcvB target sites within the 5' UTRs of all annotated *Salmonella* mRNAs.

In total, the combination of microarray analysis and prediction of C/A-rich target sites led to the identification of $>$30 potential GcvB consensus R1-dependent targets, whereof at least ten could be validated by Western blots or FACS (fluorescence activated cell sorting) analysis of GFP reporter-gene fusions. Regulation of one target, *cycA* mRNA, turned out to be independent of the G/U-rich consensus R1, which is strictly required for regulation of all other identified targets. Because all currently known targets are amino acid or peptide transporters or genes involved in amino acid biosynthesis, it is speculated that GcvB RNA has a global role in amino acid metabolism.

4.1. Results

4.1.1. Expression of GcvB from an arabinose inducible plasmid

To identify additional GcvB targets besides the known ABC transporter mRNAs, an sRNA pulse-overexpression strategy similar to previously described experimental target identification approaches was taken (see Section 2.4.1.2). For this purpose, plasmid pBAD-GcvB (pKP1-1) was used which expresses GcvB wild-type RNA under control of an arabinose-inducible P_{BAD} promoter. In addition, plasmids pBAD-GcvB$_{\Delta R1}$ (pKP2-6) and pBAD-GcvB$_{\Delta R2}$ (pKP30-1) were included for expression of GcvB$_{\Delta R1}$ and GcvB$_{\Delta R2}$ mutant RNAs, in which either the highly conserved G/U rich region R1 or consensus R2 are deleted (Fig. 4.1, left panel). To confirm inducible expression of these sRNAs, *Salmonella* wild-type carrying a pBAD control plasmid (pKP8-35) and a *Salmonella gcvB* mutant carrying either pBAD control (C), pBAD-GcvB wild-type (GcvB), pBAD-GcvB$_{\Delta R1}$ (ΔR1), or pBAD-GcvB$_{\Delta R2}$ (ΔR2) plasmid were grown to mid exponential phase (OD_{600} 1.0) and treated with L-arabinose (0.2% final concentration) for up to 15 min. RNA samples were taken prior to and at different time-points after induction (2.5', 5', 10', 15'). Expression of GcvB wild-type and the mutant RNAs was confirmed on Northern Blots and they accumulated to comparable levels after, *e. g.*, 10 min of induction (see Figure 4.1). Chromosomal GcvB shows an $\approx$2-fold increase after arabinose addition; however, induction of GcvB RNAs from the plasmids lead to $\approx$2.5-fold higher levels than the chromosomal GcvB after 10 min.

4.1.2. Microarray-based identification of GcvB target mRNAs in *Salmonella*

The same strains as described above were grown to an OD_{600} of 1.0, and RNA samples were taken after 10 min of L-arabinose induction (0.2% final concentration). Subsequently, mRNA changes upon sRNA induction were monitored on whole-genome *S. typhimurium* microarrays. Figure 4.2 shows a venn diagram for the different microarray sets with the number of genes that show $\geq$2-fold changes and a p-value $\leq$0.01 when compared to *Salmonella* Δ*gcvB* carrying the pP_{BAD} control plasmid. Of the 4716 *S. typhimurium* open reading frames represented on the microarrays, multiple transcripts were altered more than 2-fold in *Salmonella* wild-type compared to the Δ*gcvB* mutant

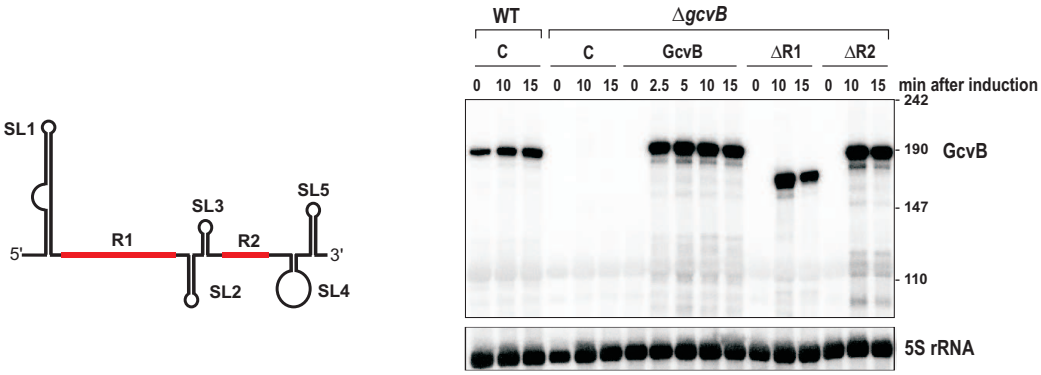


Figure 4.1: P_{BAD}-inducible expression of GcvB variants. GcvB wild-type (GcvB) and $GcvB_{\Delta R1}$ (ΔR1, deletion of residues 66-89) or $GcvB_{\Delta R2}$ (ΔR2, deletion of residues 136-144) mutant RNAs were expressed in *Salmonella* SL1344 Δ*gcvB* from the P_{BAD}-inducible promoter by addition of L-arabinose (0.2% final concentration) to cultures grown to an OD_{600} of 1.0. The highly conserved regions, R1 and R2, that are deleted in the mutant RNAs are shown in red in the simplified GcvB structure on the left. Samples were taken prior to (0 min) and at the indicated time-points (2.5, 5, 10, and 15 min) after induction. Northern Blot analysis indicates strong induction of GcvB wild-type and mutant RNAs after 10 minutes of induction. Samples taken from *Salmonella* wild-type harbouring a pBAD control plasmid (C) revealed that chromosomal GcvB is $\approx$ 2-fold increased after 10 min of arabinose addition.

strain or upon GcvB pulse-expression (see Table 4.1). The known GcvB targets *oppA* and *dppA* were repressed in the three sets 'WT + C', 'Δ*gcvB* + GcvB', and 'Δ*gcvB* + ΔR2' (Table 4.1, genes indicated in blue). As expected, they are not repressed upon pulse-overexpression of $GcvB_{\Delta R1}$ mutant RNA, as consensus R1 is strictly required for binding of these targets (Table 4.1, column 'Δ*gcvB* + ΔR1'). Also *gltI* mRNA is downregulated more than 2-fold in two sets ('Δ*gcvB* + GcvB' and 'Δ*gcvB* + ΔR2'), but did not pass the p-value in the set 'WT + C' (Table 4.1, genes regulated in two sets are indicated in red). These three known targets all encode the first genes of the operons *oppABCDF*, *dppABCDF*, and *gltIJKL*, respectively. Also the downstream genes of *oppA*, *dppA*, and *gltI* are significantly downregulated in the microarray experiments (see Table 4.1), indicating that GcvB probably reduces expression of the whole polycistronic mRNAs of these operons. For all of these genes, relatively consistent fold-changes were observed after pulse-expression of GcvB wild-type and the ΔR2 mutant RNA (Table 4.1, compare columns 'ΔgcvB + GcvB' and 'ΔgcvB + ΔR2'). Only *gltI* mRNA showed $\approx$ 2-fold higher downregulation upon pulse-expression of the $GcvB_{\Delta R2}$ mutant RNA. The previously characterized targets *livJ*, *livK*, STM4351, and *argT* are not included in any of the venn diagram groups, as they show less than 2-fold changes in all microarray sets, which was set as the threshold here. For example, *argT* and STM4351 are downregulated 1.5-fold and 1.9-fold, respectively, upon GcvB wild-type pulse-expression (see Table 4.1, genes indicated in grey). Almost no change is observed for these genes between the wild-type and *gcvB* deletion strain (Table 4.1, column 'WT + C'). This indicates that additional, so far unknown GcvB targets could also show only slight mRNA changes and will be missed due to setting the 2-fold change threshold. Furthermore, genes that are not expressed at all under the examined growth condition will be missed in this approach.

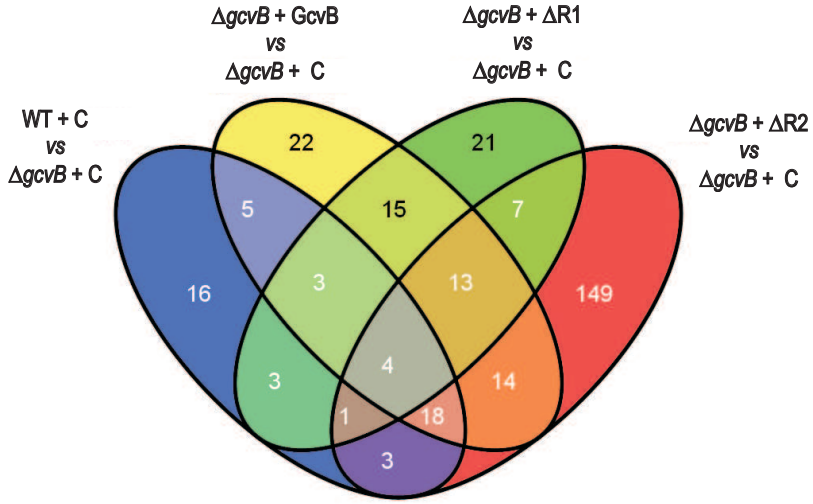


Figure 4.2: Overlap between different microarray experiments. Venn diagram showing the overlap of genes that are regulated more than 2-fold (p-value $\leq$0.01) in different microarray experiments of GcvB pulse-overexpression. The number of genes changed in *Salmonella* wild-type harbouring the pBAD control plasmid (C) or *Salmonella* Δ*gcvB* carrying the pBAD-GcvB wild-type (GcvB), pBAD-GcvB$_{\Delta R1}$ (ΔR1), or pBAD-GcvB$_{\Delta R2}$ (ΔR2) compared to the expression in *Salmonella* Δ*gcvB* carrying the pBAD-control plasmid (C) are shown. The venn diagram was generated by VENNY (bioinfogp.cnb.csic.es/tools/venny/index.html) using the names of the genes which are changed in the different experiments as input sets.

Besides the *oppABCDF*, *dppABCDF*, and *gltIJKL* operons, harbouring the known direct targets, *oppA*, *dppA*, and *gltI*, additional genes that are repressed by GcvB wild-type or GcvB$_{\Delta R2}$ RNA but not by GcvB$_{\Delta R1}$ were identified: *gdhA*, *ydgR (tppB)*, *ygjU*, *yhjG* (Table 4.1, genes indicated in blue are repressed in three sets), and *acs*, *asd*, *lrp*, *rplT*, *rpmE*, *rpmI*, *yaeC*, *ydiJ*, *yfiD*, *yggH*, STM1368, and STM3333 (Table 4.1, genes indicated in red are repressed in two sets). Of these, *ydgR*, *ygjU*, and *yaeC* encode for amino acid transport proteins, whereas *gdhA* and *asd* encode for proteins involved in biosynthesis of several amino acids, and *lrp* for a transcriptional regulator, respectively. The Lrp protein belongs to the AsnC family of transcriptional regulators and affects expression of diverse operons (Hung *et al.*, 2002). It controls, for example, several genes involved in the high-affinity branched-chain amino acid transport system and is a mediator of the leucine response in *E. coli* (Haney *et al.*, 1992; Willins *et al.*, 1991).

To confirm regulation of these genes, translational *gfp* fusions were constructed as described in Table 8.13 and Table 8.14 under Materials and Methods. The different *gfp* fusions contain between seven and 19 amino acids of the N-termini of these target genes and 5' UTR parts that start at transcriptional start sites derived from one of the following sources: own 5' RACE results, promoters described in the EcoCyc[1] database or the literature, or deep-sequencing results of Hfq-bound RNA in *Salmonella* (Sittka *et al.*, 2008).

[1] www.ecocyc.org/

Table 4.1: Genes which show more than 2-fold change in at least two GcvB microarray experiments and previously known target genes.

The last four columns indicate the fold-changes for the genes given in the first column in the four different microarray experiments. The values correspond to the fold-changes of a gene in the indicated strains compared to the expression level in the *Salmonella* Δ*gcvB* strain harbouring the pBAD-control plasmid (set 'Δ*gcvB* + C'). A negative value corresponds to downregulation and a positive value to upregulation of the mRNA, respectively. Brackets indicate fold-changes ≥2.0 which did not satisfy the p-value threshold of 0.01; fold-changes ≥2-fold are set in bold. The known GcvB targets from the previous section (*argT, livJ, livK*, and STM4351) which did not satisfy the 2-fold change or p-value threshold, are listed in grey. Genes that show more than 2-fold change (p-value < 0.01) in all experiments are indicated in green, whereas those with a significant ≥2-fold change in the three sets WT + C, Δ*gcvB* + GcvB, and Δ*gcvB* + ΔR2 or in the two sets Δ*gcvB* + GcvB and Δ*gcvB* + ΔR2 are indicated in blue and red, respectively.

Gene	**Description**	WT + C	Δ*gcvB* + GcvB	Δ*gcvB* + ΔR1	Δ*gcvB* + ΔR2
acs	acetyl-CoA synthetase	-1.5	**-3.4**	+1.2	**-5.2**
argT	ABC superfamily (bind-prot), lysine/arginine/ornithine transport protein	-1.3	-1.5	±1.0	-1.5
asd	aspartate-semialdehyde dehydrogenase	+1.0	**-2.1**	+1.3	**-2.0**
bfr	bacterioferrin, iron storage homoprotein	+1.6	**+2.1**	**+2.2**	+1.3
caiF	transcriptional regulator of *cai* and *fix* operon	**+2.7**	(+3.3)	**+3.2**	**+2.3**
carA	carbamoyl-phosphate synthetase, glutamine-hydrolysing small subunit	(+6.3)	**+7.8**	**+7.8**	**+10.2**
cycA	APC family, D-alanine/D-serine/glycine transport protein	**-3.4**	**-5.1**	**-3.6**	**-6.0**
dppA	ABC superfamily (peri-perm), dipeptide transport protein	**-6.8**	**-13.2**	-1.7	**-12.9**
dppB	ABC superfamily (membrane), dipeptide transport protein 1	**-2.1**	**-3.1**	-1.2	**-3.0**
dppC	ABC superfamily (membrane), dipeptide transport protein 2	**-2.6**	**-3.2**	-1.2	**-3.5**
dppD	ABC superfamily (ATP-bind.), dipeptide transport protein	**-3.8**	**-6.9**	-1.3	**-9.1**
dppF	ABC superfamily (ATP-bind.), dipeptide transport protein	**-2.9**	**-4.0**	-1.3	**-4.1**
dps	stress response DNA-binding protein; starvation induced resistance to H_2O_2	+1.8	**+2.1**	**+2.2**	+1.4
elaB	putative inner membrane protein	**+2.2**	**+2.3**	**+2.4**	+1.5
fbaB	3-oxoacyl-[acyl-carrier-protein] synthase I	+1.8	**+2.1**	**+2.2**	+1.2
fimF	putative fimbrial protein	**-2.7**	-1.2	+1.4	**-2.1**
fimI	fimbrial protein internal segment	**-4.4**	-1.1	+1.7	**-2.6**
flgB	flagellar biosynthesis, cell-proximal portion of basal-body rod	**+2.5**	**+2.4**	+1.8	+1.8
flgD	flagellar biosynthesis, initiation of hook assembly	**+2.5**	**+2.0**	+1.7	+1.6
gapA	glyceraldehyde-3-phosphate dehydrogenase A	+1.5	**+2.1**	**+2.1**	+1.6
garL	2-dehydro-3-deoxy-galactarate aldolase	**-2.1**	**-2.1**	-1.9	-1.8
gdhA	glutamate dehydrogenase, NADP-specific	**-2.3**	**-3.4**	-1.8	**-3.2**
gltI	ABC superfamily (bind-prot), glutamate/aspartate transporter	(-4.2)	**-8.5**	-1.1	**-14.8**
gltJ	ABC superfamily (membrane), glutamate/aspartate transporter	**-2.7**	**-3.1**	-1.4	**-3.5**
gltK	ABC superfamily (membrane), glutamate/aspartate transporter	**-2.7**	**-4.2**	-1.6	**-4.7**

continued on next page

Gene	Description	WT + C	Δ*gcvB* + GcvB	Δ*gcvB* + ΔR1	Δ*gcvB* + ΔR2
gltL	ABC superfamily (ATP-bind.), glutamate/aspartate transporter	(-3.0)	**-5.6**	-1.7	**-6.2**
glyS	glycine tRNA synthetase, beta subunit	(-2.6)	(-2.8)	**-2.5**	**-2.4**
grxB	glutaredoxin 2	(+2.4)	**+2.8**	**+2.7**	**+2.5**
hepA	RNA polymerase associated protein, putative SNF2 family RNA helicase	(-2.0)	**-2.4**	**-2.2**	**-2.7**
katE	catalase; hydroperoxidase HPII(III), RpoS dependent	(+2.1)	**+2.2**	**+2.3**	+1.3
livJ	ABC superfamily (bind-prot), branched-chain amino acid transporter, high-affinity	±1.0	-1.6	-1.2	-1.4
livK	ABC superfamily (bind-prot), branched-chain amino acid transporter, high-affinity	-1.1	-1.2	±1.0	-1.1
lrp	regulator for *lrp* regulon and high-affinity branched-chain amino acid transport system; mediator of of leucine response (AsnC family)	-1.2	**-3.1**	(-2.1)	**-2.1**
miaE	hydroxylase for synthesis of 2-methylthio-cis-ribozeatin in tRNA	(+8.3)	-1.1	**+8.2**	**+11.8**
msyB	acidic protein suppresses mutants lacking function of protein export	(+2.3)	**+3.0**	**+3.4**	+1.8
ompW	outer membrane protein W; colicin S4 receptor; putative transporter	**+2.4**	(+2.6)	**+3.6**	+1.5
oppA	ABC superfamily (periplasm), oligopeptide transport protein with chaperone properties	**-7.3**	**-18**	-1.4	**-23.2**
oppB	ABC superfamily (membrane), oligopeptide transport protein	**-3.3**	**-6.3**	-1.3	**-7.2**
oppD	ABC superfamily (ATP-bind.), oligopeptide transport protein	**-2.9**	**-4.7**	-1.5	**-5.8**
oppF	ABC superfamily (ATP-bind.), oligopeptide transport protein	**-2.4**	**-2.8**	-1.3	**-3.9**
phoL	putative phosphate starvation-inducible protein, ATP-binding	+1.9	**+2.9**	**+2.4**	+1.7
pipC	pathogenicity island encoded protein: homologous to ipgE of Shigella	**+2.4**	**+2.2**	+1.9	+1.3
ptsG	sugar specific PTS family, glucose-specific IIBC component	**-2.6**	(-2.1)	**-2.1**	-1.9
pyrB	aspartate carbamoyltransferase, catalytic subunit	(+65.5)	**+74.6**	**+68.4**	**+89.8**
pyrE	orotate phosphoribosyltransferase	(+5.7)	**+7.1**	**+6.7**	**+7.7**
pyrI	aspartate carbamoyltransferase, regulatory subunit (allosteric regulation)	(+52.8)	**+56.8**	**+47.2**	**+52.2**
pyrL	*pyrBI* operon leader peptide	**+2.3**	**+2.6**	(+2.6)	**+2.8**
rnpA	RNase P, protein component (protein C5), processes tRNA, 4.5S RNA	(-2.2)	**-2.8**	**-2.9**	**-2.6**
rplT	50S ribosomal subunit protein L20	(-2.6)	**-2.9**	(-2.1)	**-2.7**
rpmE	50S ribosomal subunit protein L31	(-2.3)	**-2.0**	-1.9	**-2.2**
rpmI	50S ribosomal subunit protein L35	(-2.1)	**-2.5**	-1.8	**-2.3**
rpoB	RNA polymerase, beta subunit	(-2.4)	(-2.7)	**-2.5**	**-2.6**
rpsL	30S ribosomal subunit protein S12	(-2.0)	**-2.8**	**-2.3**	**-2.4**
rtsA	putative AraC-type DNA-binding domain-containing protein	**+3.7**	**+3.3**	**+2.7**	**+2.6**
rtsB	putative bacterial regulatory proteins, *luxR* family	(+3.7)	(+2.9)	**+2.6**	**+2.5**
ssaJ	Secretion system apparatus: homology with the *yscJ/mxiJ/prgK* family of lipoproteins	**-3.1**	+1.2	-1.1	**-2.6**
uraA	NCS2 family, uracil transport protein	(+2.8)	(+3.4)	**+4.7**	**+2.3**
wraB	*trp*-repressor binding protein	+1.7	**+2.1**	**+2.4**	+1.3

continued on next page

Gene	Description	WT + C	Δ*gcvB* + GcvB	Δ*gcvB* + ΔR1	Δ*gcvB* + ΔR2
yaeC	putative outer membrane lipoprotein, D-methionine transport system substrate-binding protein (*metQ*)	-1.8	**-2.4**	-1.4	**-2.2**
ybbN	putative thioredoxin protein	+2.0	**+3.5**	**+3**	+1.1
yccJ	putative cytoplasmic protein	(+2.3)	**+3.0**	**+3.0**	+1.8
ydgR	putative POT family, peptide transport protein (*tppB*)	**-2.2**	**-3.7**	-1.4	**-3.8**
ydiJ	putative oxidase	-1.9	**-3.0**	-1.3	**-3.1**
yeeF	putative APC family, amino acid transport protein	(-3.7)	**-3.0**	**-2.7**	**-2.5**
yegQ	putative protease	(-2.6)	**-3.1**	**-3.0**	**-2.2**
yfiD	putative formate acetyltransferase	**+2.4**	**+4.3**	(+4.4)	**+4.9**
yggH	putative S-adenosylmethionine-dependent methyltransferase	-1.8	**-2.2**	-1.7	**-2.1**
ygjU	(*sstT*), serine/threonine transporter SstT, putative dicarboxylate permease	(-4.2)	**-6.4**	+1.2	**-5.5**
yhjG	putative inner membrane protein	**-6.1**	**-14.1**	-1.5	**-14.1**
yidC	putative preprotein translocase subunit YidC	(-2.2)	(-2.4)	**-2.2**	**-2.4**
yjgF	putative translation initiation inhibitor	**+4.4**	**+6.1**	**+6.5**	**+5.4**
ymdA	putative periplasmic protein	**+2.1**	**+2.0**	+1.7	+1.7
SL1344-0022	SL1344 specific protein	+1.7	**+2.0**	**+2.0**	(+2.0)
SL1344-0031	SL1344 specific protein	+1.9	**+2.9**	**+2.5**	(+2.4)
STM1055	Gifsy-2 prophage	**+2.1**	**+2.1**	**+2.0**	**+2.2**
STM1368	putative Na^+-dicarboxylate symporter	+1.8	**+2.5**	(+2.1)	**+2.2**
STM1747[1]	putative inner membrane protein	**-3.0**	**-4.3**	-1.3	**-5.0**
STM2746	putative excinuclease ATPase subunit	+1.8	**+2.4**	**+2.4**	**+2.5**
STM2747	putative cytoplasmic protein	+1.9	**+2.6**	**+2.6**	**+2.7**
STM3333	putative purine-cytosine permease (*codB*)	(+4.5)	**+5.6**	(+5.4)	**+7.5**
STM3334	putative cytosine deaminase	(+4.6)	**+4.9**	**+5.2**	**+4.8**
STM3841	putative inner membrane protein	(-2.0)	(-2.6)	**-2.6**	**-2.8**
STM4313	putative cytoplasmic protein	**+2.9**	(+2.4)	**+2.1**	+1.8
STM4351	putative arginine-binding periplasmic protein	-1.6	-1.9	-1.3	-1.8
STM4535	putative PTS permease	(-2.1)	**-2.3**	**-2.6**	-1.4
STM4537	putative PTS permease	**-2.8**	**-2.7**	**-2.9**	-1.8
STM4538	putative PTS permease	-1.8	**-2.3**	**-2.5**	-1.5
STM4539	putative glucosamine-fructose-6-phosphate aminotransferase	**-2.8**	**-2.8**	**-3.0**	-1.7
STM4540	putative glucosamine-fructose-6-phosphate aminotransferase	(-3.5)	**-2.6**	**-3.1**	-1.6

The resulting fusion plasmids and control plasmid pXG-1 (*gfp*) were each combined with either pP_L*gcvB*, pP_L*gcvB*$_{\Delta R1}$, pP_L*gcvB*$_{\Delta R2}$, which express the sRNAs from a constitutive $P_{LlacO-1}$ promoter, or control vector pTP011 in an *E. coli gcvB* deletion strain as described in the previous section and in Sharma *et al.* (2007). Figure 4.3A shows reduced colony fluorescence monitored on agar-plates for the *ygjU, yaeC, tppB (ydgR), gdhA*, and *lrp* fusions in the strains harbouring pP_L*gcvB* or pP_L*gcvB*$_{\Delta R2}$ similar to the pattern observed for the known target *oppA*. The *asd::gfp* fusion was neither regulated on agar plates, nor on Western Blots and in FACS analyses of GFP fusion protein levels (Fig. 4.3). In contrast, a consensus R1 dependent regulation could be confirmed for the *ygjU::gfp* and *gdhA::gfp* fusions on Western blots (Fig. 4.3B) and by FACS analysis (Fig. 4.3C). No

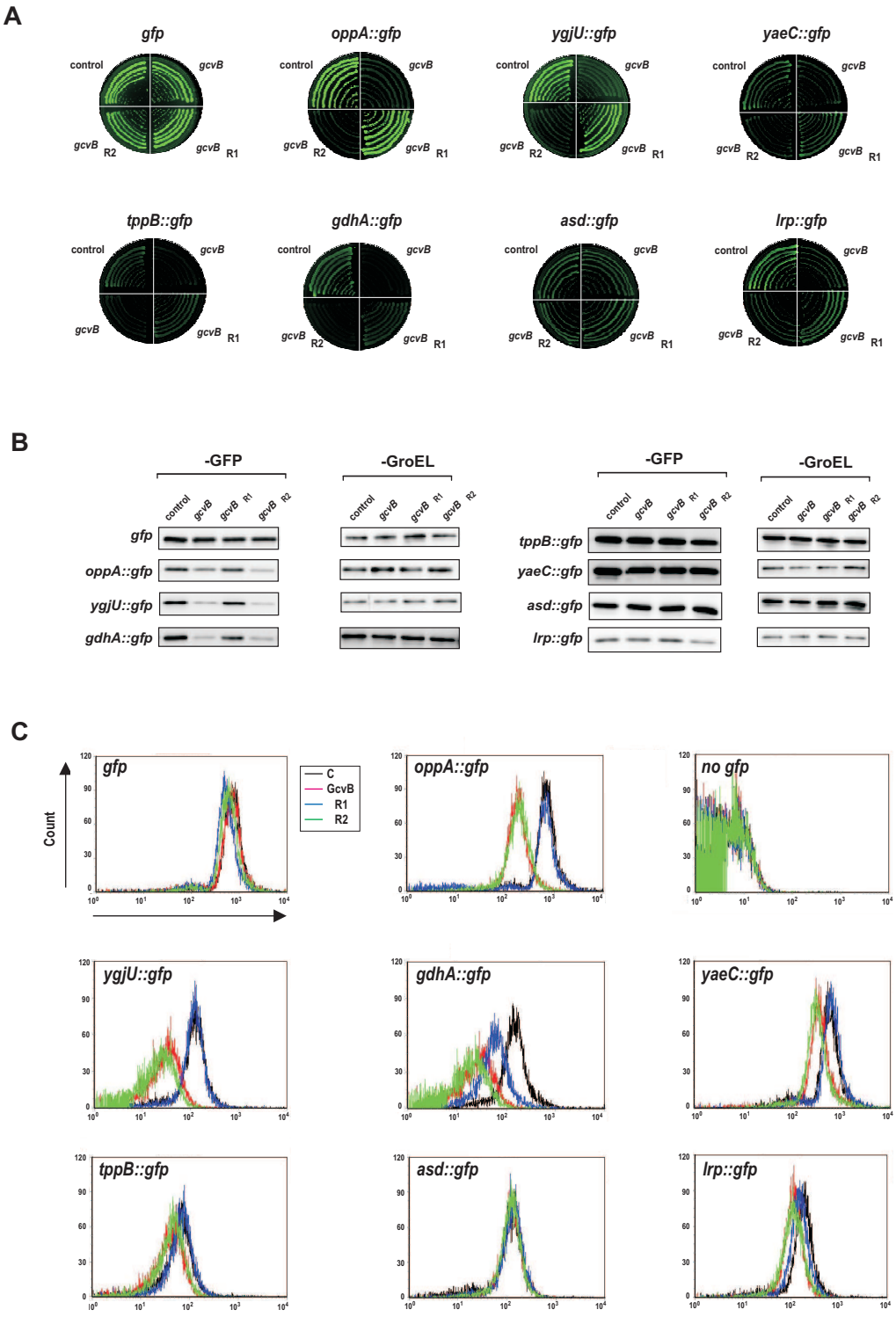
[1] STM17147 is located antisense to the *oppA* 5' UTR. Thus, as due to cDNA construction no strand differentiation is possible it is unclear whether repression is a direct effect of GcvB or of co-hybridization with *oppA* cDNA.

clear regulation as previously observed on agar plates could be confirmed for *tppB::gfp*, *yaeC::gfp*, and *lrp::gfp* on Western Blots, however, FACS analysis indicated regulation by GcvB for *yaeC::gfp*. Moreover, an $\approx$3-fold increase of chromosomal *lrp* mRNA level was observed on Northern blots for a *Salmonella gcvB* deletion strain at mid-log growth compared to the isogenic wild-type strain (data not shown), indicating a direct or indirect influence by GcvB RNA consistent with the microarray results.

4.1.3. Prediction of C/A elements reveals further targets

In Chapter 3 it was shown, that GcvB targets C/A-rich regions in the 5' UTR of seven ABC transporter RNAs. Also the validated new GcvB targets, *ygjU*, *yaeC*, and *gdhA* are predicted to be bound by GcvB consensus R1 sequence at C/A-rich sites within their 5' UTRs (Fig. 4.4A, left). Similarily, *asd* and *ydgR* contain potential C/A-rich interaction sites that could interact with GcvB RNA (Fig. 4.4A, right). This extended target set including the new microarray targets was used to define a consensus motif for the GcvB target site. The *gfp* fusion sequences of the seven previously known targets (*dppA*, *oppA*, *livJ*, *livK*, *argT*, STM4351, and *gltI*), the six R1 dependent (*ydgR*, *ygjU*, *yaeC*, *gdhA*, *asd*, and *lrp*) targets from the microarray analysis, as well *cycA*, which was repressed in all microarray sets (Table 4.1, indicated in green) and will be discussed in more detail below (Section 4.1.5), were used as input for `MEME` motif identification (Bailey *et al.*, 2006) with the following parameters: number of different motifs: 10; minimum number of sites: 8; maximum number of sites: 14; minimum motif width: 6; maximum motif width: 25. This allows the identification of up to ten motifs with a length between 6 and 25 bp within the input sequences. The upper size restriction of 25 nt was selected to cover long interaction sites in the size range of the 24-nt long G/U rich R1 linker sequence. Furthermore, the motif has to be present in at least 50% of the 14 input sequences as defined by the parameters for minimum and maximum number of sites.

Figure 4.3 *(facing page)*: New targets identified on microarrays upon GcvB pulse-expression. (A) Agar plate-based assay of colony fluorescence of *E. coli* Δ*gcvB recA*$^-$ strains carrying control plasmid pXG-1 which expresses full-length *gfp*, the *oppA::gfp* (pJL19-1), *ygjU::gfp* (pFS133-3), *yaeC::gfp* (pFM27-1), *tppB::gfp* (pJL70-9), *gdhA::gfp* (pJL69-5), *asd::gfp* (pFS116-1), or *lrp::gfp* (pFS103-3) fusion plasmid, each in combination with control vector pTP11, or plasmids expressing *Salmonella* wild-type GcvB RNA (pP_L*gcvB*) or two of the mutant alleles (plasmids pP_L*gcvB*$_{\Delta R1}$ or pP_L*gcvB*$_{\Delta R2}$). **(B)** Western blots of target::GFP fusion proteins prepared after growth for 14 hours to stationary phase from *E. coli* Δ*gcvB recA*$^-$ carrying the indicated plasmids as in A. GroEL was probed as loading control. Fold changes of GFP fusion protein levels (upon normalization to GroEL levels) by *gcvB*, *gcvB*$_{\Delta R1}$, or *gcvB*$_{\Delta R2}$ co-expression relative to the control plasmid were: OppA::GFP, -2.6/-1.1/-3.8; YgjU::GFP, -3.6/-1.3/-7.1; GdhA::GFP, -5.3/-1.4/-4.3; TppB::GFP, +1.1/-1.1/-1.4; YaeC::GFP, 1.0/-1.1/-1.8; Asd::GFP, +1.3/+1.1/1.0; Lrp::GFP, 1.0/1.0/-1.5; GFP alone, -1.0/-2.1/-1.4. **(C)** *E. coli* Δ*gcvB recA*$^-$ strains carrying control plasmid pXG-1 (*gfp*), a non-fluorescent control plasmid pXG-0 (no gfp), the *oppA::gfp*, *ygjU::gfp*, *gdhA::gfp*, *yaeC::gfp*, *tppB::gfp*, *asd::gfp*, or *lrp::gfp* fusion plasmids in combination with control vector pTP11 (black), or plasmids expressing *Salmonella* wild-type GcvB RNA (pP_L*gcvB*, red) or two of the mutant alleles (pP_L*gcvB*$_{\Delta R1}$, blue; pP_L*gcvB*$_{\Delta R2}$, green) were grown to stationary phase and were subjected to flow cytometry analysis. All data acquired from the experiments are plotted in fluorescence histograms realized on all events measured (30,000 events). Cellular fluorescence is given in arbitrary units (GFP intensity). Regulation by GcvB wild-type or GcvB$_{\Delta R2}$ is visible as a shift of the fluorescence curves to the left to lower GFP intensities.


A
gfp
oppA::gfp
ygjU::gfp
yaeC::gfp
tppB::gfp
gdhA::gfp
asd::gfp
lrp::gfp
control
gcvB
R1
R2
B
-GFP
-GroEL
gfp
oppA::gfp
ygjU::gfp
gdhA::gfp
tppB::gfp
yaeC::gfp
asd::gfp
lrp::gfp
C
Count
C
GcvB
R1
R2
no gfp

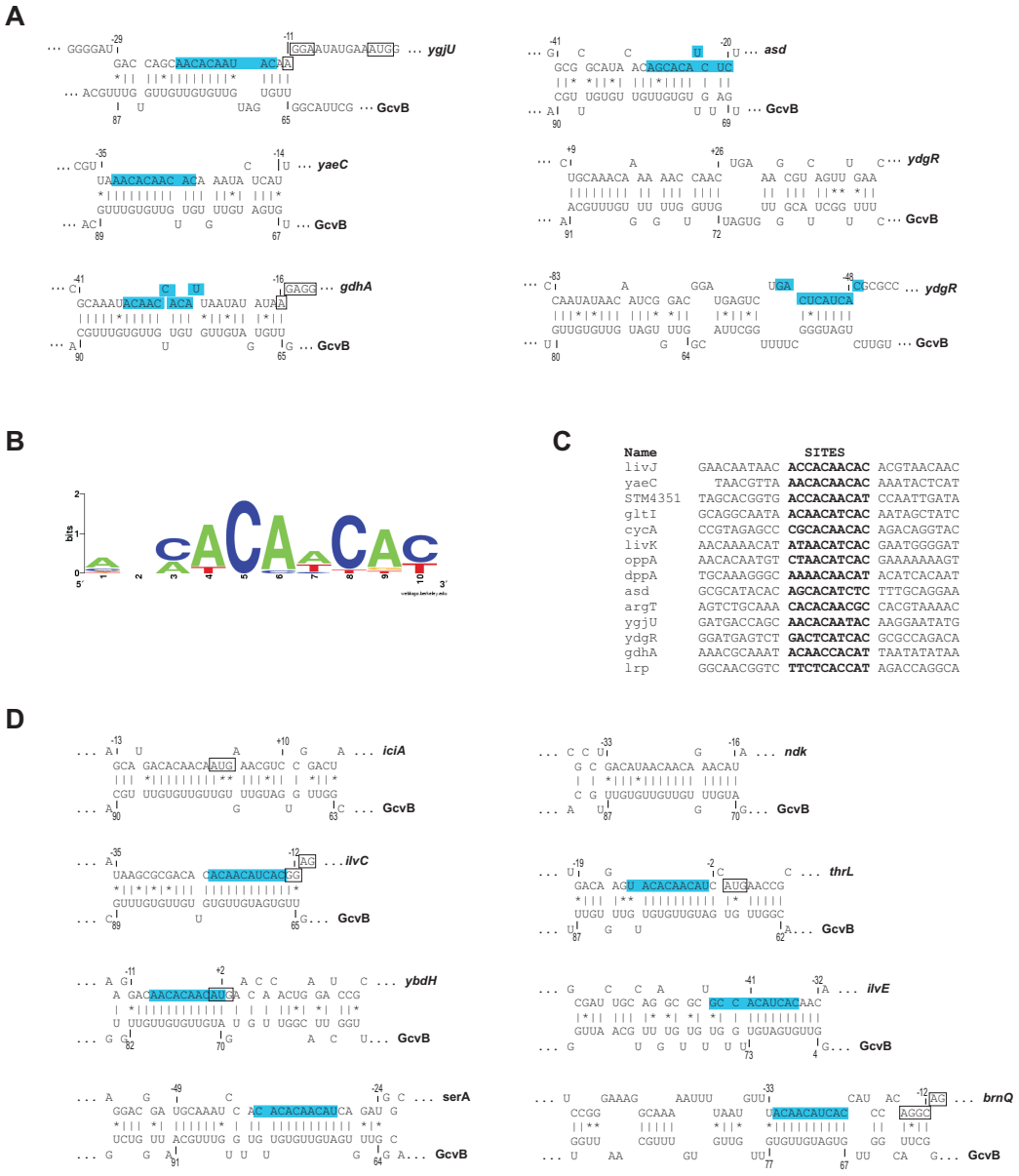


Figure 4.4: Predicted target interactions of GcvB and identification of a C/A-rich motif. **(A)** Proposed RNA duplexes formed by GcvB with five new target mRNAs (*ygjU*, *yaeC*, *gdhA*, *asd*, and *ydgR*). Positions in the target sequences are given as distance to the mRNA start codon. C/A-rich motifs are highlighted in blue in the interactions. For *ydgR* two interactions are predicted, whereof the first one is located in the coding sequence. **(B)** A consensus motif for the C/A-rich GcvB target sites was determined by `MEME` (http://meme.nbcr.net/meme4_1/cgi-bin/meme.cgi) and visualized with `WebLogo` (weblogo.berkeley.edu/). **(C)** Location of the C/A-rich motif in the input sequences. **(D)** Proposed RNA duplexes formed by GcvB with eight additional target mRNAs (*iciA*, *ilvC*, *ybdH*, *serA*, *ndk*, *thrL*, *ilvE*, and *brnQ*).

Table 4.2: Genes with predicted C/A elements and more than 2-fold change in at least one microarray experiment. All fold changes are normalized to expression in *Salmonella* Δ*gcvB* + pBAD-control. Brackets indicate fold-changes ≥2.0 which did not satisfy the p-value threshold of 0.01.

Gene	Description	WT + C	ΔgcvB + GcvB	ΔgcvB + ΔR1	ΔgcvB + ΔR2
brnQ	LIVCS family, branched chain amino acid transporter system II (LIV-II)	-1.3	**-2.0**	-1.2	(-2.0)
serA	D-3-phosphoglycerate dehydrogenase	-1.3	(-2.0)	1.0	**-2.1**
rho	transcription termination factor Rho	-1.3	-1.3	-1.1	**-2.0**
ydiJ	putative oxidase	-1.9	**-3.0**	-1.3	**-3.1**
mglC	ABC superfamily (membrane), methyl-galactoside transport protein	(-3.2)	(-4.2)	(-2.2)	**-5.0**
ilvB	acetolactate synthase I, large subunit	-1.8	(-2.4)	(-2.0)	**-2.4**
ygdL	putative enzyme	-1.6	-1.7	-1.6	**-2.8**
STM0148	putative cytoplasmic protein	-1.3	-1.3	1.2	**-3.1**
STM3334	putative cytosine deaminase	(+4.6)	**+4.9**	**+5.2**	**+4.8**
STM0293	putative cytoplasmic protein	+1.4	+1.6	+1.2	**+2.1**
flgE	flagellar biosynthesis, hook protein	**+2.1**	+1.5	+1.4	+1.4
pipC	pathogenicity island encoded protein: homologous to *ipgE* of *Shigella*	**+2.4**	**+2.2**	+1.9	+1.3
ycfR	putative outer membrane protein	-1.1	+1.0	-1.4	**+2.1**

Figure 4.4B shows the best motif identified by MEME. It is 10 nts long and present in all input sequences (Fig. 4.4C). For all input sequences, except for *lrp*, the identified motif overlaps with a mapped or proposed GcvB interaction site. For *ydgR*, the motif overlaps with the second proposed interaction in Figure. 4.4A. Thus, although all the targets are overall unrelated in sequence, they share one common motif which could be targeted by the G/U rich R1 sequence of GcvB RNA and thereby mediate regulation of these targets. However, direct binding has to be confirmed experimentally in future work either by compensatory base-pair exchanges or *in vitro* footprinting analyses of GcvB-target mRNA complexes.

4.1.4. Identification of additional mRNAs that contain the GcvB target site

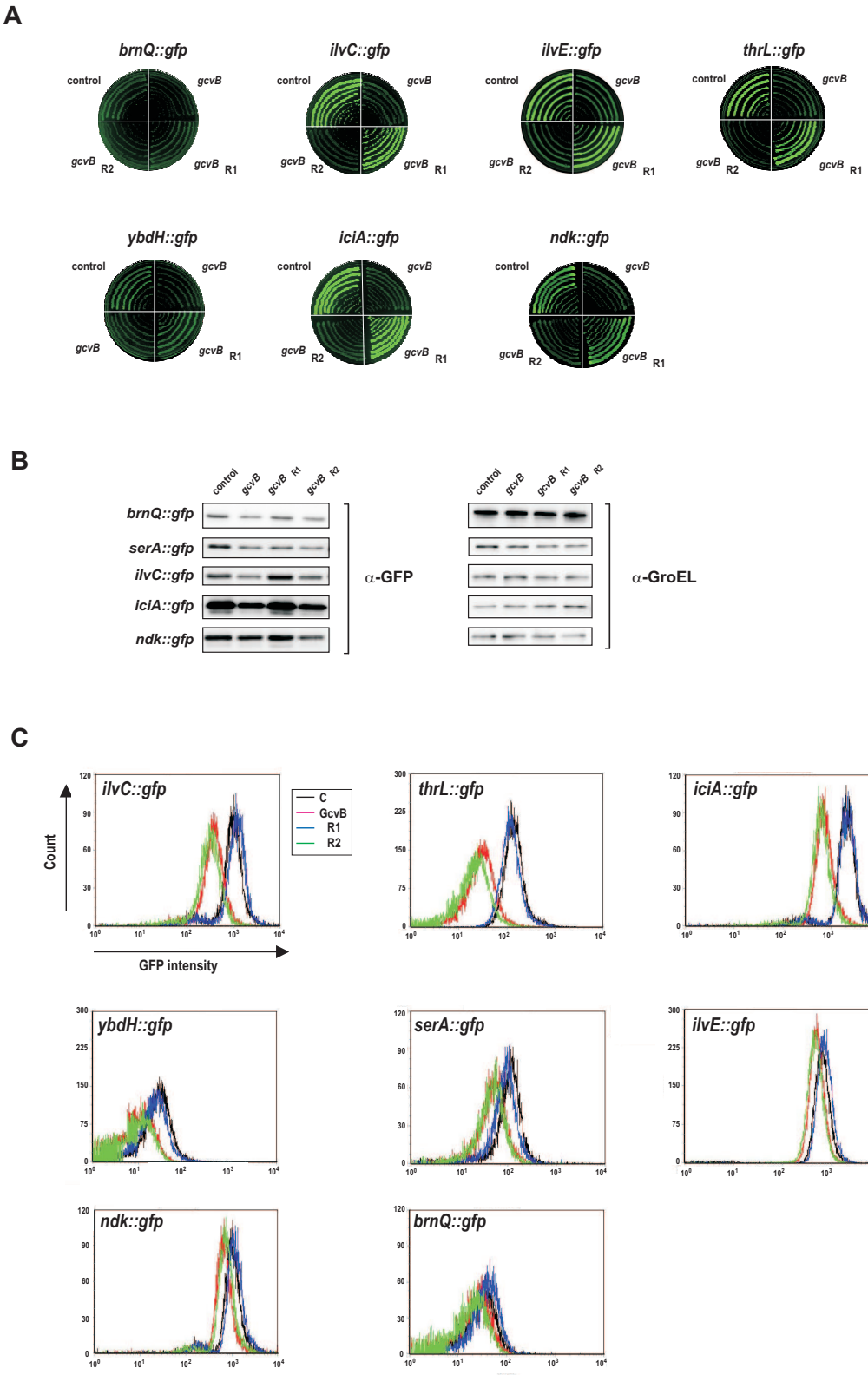
Altough several new GcvB targets could be identified based on the microarray approach, some further targets could have been missed due to low or no expression under the examined growth condition. Moreover, some targets could not have reached the 2-fold change or significance thresholds like the previously characterized targets *livJ, livK*, STM4351, and *argT* which show less than 2-fold downregulation. In contrast, bioinformatics-based predictions should in principle be able to predict all targets; however, they often predict also a high number of false positives. To refine a bioinformatics-based target identification, the search was restricted to the identification of additional R1-dependent GcvB targets involving a C/A-rich target site. To see if the 10-nt long motif which was identified in the previous targets (Fig. 4.4B) can be found in additional mRNAs and might reveal further GcvB targets, MAST (Bailey & Gribskov, 1998) searches were performed with the corresponding position-specific weight matrix for the C/A-rich GcvB target site in a database composed of the sequences -70 to +30 (according to the start codon) of all annotated *Salmonella* ORFs. In this set of 4424 sequences, the motif could be identified in 245 sequences with an E-value

< 200. With this E-value threshold, all 14 input sequences, except for *lrp*, are found in the list of the 245 mRNAs (data not shown). To narrow down the set of additional GcvB targets for experimental verification, it was examined which of these mRNAs are also regulated in at least one of the microarray datasets. The mRNAs of *brnQ, serA, rho, ydiJ, mglC, ilvB, ygdL*, and STM0148 are downregulated, whereas STM3334, STM0293, *flgE, pipC, ycfR* are upregulated by GcvB in at least one microarray set (see Table 4.2).

Additionally, potential target interactions were predicted using `RNAhybrid` (Rehmsmeier *et al.*, 2004) between the regions -70/+30 of these 245 mRNAs and an extended GcvB region R1 (the 24-nt long R1 region plus 12 and 13 nt flanking regions) with two parameter sets, one without limitations for bulge/internal loops and another where these loops were restricted to a length of one nucleotide. After removal of interactions that had no helix with at least nine consecutive base-pairs, the lists contained 88 predicted GcvB-target interactions without limitations for loops and 91 interactions with loops of maximal one nucleotide, respectively. Helices of nine base-pairs were observed as some of the shortest biochemically-validated interactions, *e. g.*, in case of OxyS:*fhlA* and RyhB:*sodB* (Argaman & Altuvia, 2000; Geissmann & Touati, 2004). In this step, *livJ, asd*, and *ydgR* were sorted out. With both parameter sets, the best interaction in terms of free energy is the one with *dppA* mRNA which was also mapped experimentally in the previous Chapter. In both sets of predicted interactions, the second ranked target is *ilvC*, which encodes for an acetohydroxy acid isomeroreductase and is involved in isoleucine and valine biosynthesis. However, this target was not regulated in the microarray list.

To further narrow down the lists, targets were selected from the predictions which showed strong interactions (no more than one G:U basepair in the middle of the longest helix, except for the known targets) leading to 35 and 33 final interactions. These included nine (*dppA, gltI, argT, oppA, livK, gdhA*, STM4351, *yaeC*, and *ygjU*) and eight (*dppA, gltI, argT, oppA, livK, gdhA*, STM4351, and *cycA*) of the input targets, respectively, that were used for the motif identification. Interactions for *ygjU, yaeC*, and *cycA* were included in only one of each set, as the different loop parameters sometimes lead to differences in the predicted interactions. For example, the longest helix in the

Figure 4.5 *(facing page)*: New GcvB targets based on biocomputational predicions. (A) Agar plate-based assay of colony fluorescence of *E. coli* Δ*gcvB recA*$^-$ strains carrying the same sRNA plasmids like in Figure 4.3A in combination with the *brnQ::gfp* (pFS105-3), *ilvC::gfp* (pJL68-1), *ilvE::gfp* (pSP25-7), *thrL::gfp* (pSP20-1), *ybdH::gfp* (pSP21-2), *iciA::gfp* (pFS121-1), or *ndk::gfp* (pFS115-2) fusion plasmid. **(B)** Western blots of target::GFP fusion proteins prepared after growth to stationary phase from *E. coli* Δ*gcvB recA*$^-$ carrying the indicated plasmids as in **A** or a *serA::gfp* fusion plasmid (pFS117-1). GroEL was probed as loading control. Fold changes of GFP fusion protein levels (upon normalization to GroEL levels) by *gcvB*, *gcvB*$_{\Delta R1}$, or *gcvB*$_{\Delta R2}$ co-expression relative to the control plasmid were: BrnQ::GFP, -2.0/-1.1/-2.0; SerA::GFP, -1.4/+1.1/-1.2; IlvC::GFP, -1.9/+1.4/-1.2; IciA::GFP, -2.6/-1.7/-3.6; and Ndk::GFP, -1.4/+1.2/-1.1. **(C)** *E. coli* Δ*gcvB recA*$^-$ strains carrying the *ilvC::gfp, thrL::gfp, iciA::gfp, ybdH::gfp, serA::gfp, ilvE::gfp, ndk::gfp*, or *brnQ::gfp* fusion plasmids in combination with control vector pTP11 (black), or plasmids expressing *Salmonella* wild-type GcvB RNA (pP_L*gcvB*, red) or two of the mutant alleles (pP_L*gcvB*$_{\Delta R1}$, blue; pP_L*gcvB*$_{\Delta R2}$, green) were grown to stationary phase and were subjected to flow cytometry analysis. All data acquired from the experiments are plotted in fluorescence histograms realized on all events measured (30,000 events). Cellular fluorescence is given in arbitrary units (GFP intensity). Regulation by GcvB wild-type or GcvB$_{\Delta R2}$ is visible as a shift of the fluorescence curves to the left to lower GFP intensities.


A
brnQ::gfp
ilvC::gfp
ilvE::gfp
thrL::gfp
ybdH::gfp
iciA::gfp
ndk::gfp
control
gcvB
R1
R2
B
brnQ::gfp
serA::gfp
ilvC::gfp
iciA::gfp
ndk::gfp
α-GFP
α-GroEL
C
Count
GFP intensity
C
GcvB
R1
R2

interaction predicted for *cycA* with no restrictions for bulge loops is only eight base-pairs long and, hence, was rejected in the step in which interactions that had no helix with at least nine subsequent base-pairs were sorted out. In contrast, limitation of the loop length to one nucleotide leads to the prediction of a different interaction which satisfies the nine base-pair threshold and is thus kept in the list of final interactions.

Five genes (*brnQ, serA, rho, mglC,* and *ilvB*) of the set of targets with a predicted C/A-rich element which are downregulated in at least one microarray experiment (Table 4.2) were also included in the final lists of predicted interactions. Translational *gfp* fusions were constructed for these five genes, and in addition *ilvC, ilvE, thrL, ybdH, sbp, mltC, yieG,* and STM2179 were selected from the interaction list for construction of GFP fusions. Additional predictions for interactions of GcvB R1 with the -70/+30 regions of all annotated *Salmonella* ORFs indicated that also *iciA* and *ndk* could form long duplexes with GcvB RNA (see Fig. 4.4D). Therefore these two genes were also selected for construction of GFP fusions.

Only seven of the new fusions, *brnQ::gfp, ilvC::gfp, ilvE::gfp, thrL::gfp, ybdH::gfp, iciA::gfp* and *ndk::gfp,* displayed sufficient GFP levels to monitor colony fluorescence on agar plates and indicated an R1-dependent repression of these genes (Fig. 4.5A). Regulation of *brnQ::gfp, ilvC::gfp* and *iciA::gfp* could be confirmed on Western blots (Fig. 4.5B). On the Western blot in Figure 4.5B the *ndk::gfp* fusion seems to be strongly downregulated by $GcvB_{\Delta R2}$. However, normalization to GroEL levels indicates a loading artefact as also a lower GroEL amount was detected in this lane. FACS analysis showed a strong R1-dependent regulation for *ilvC::gfp, thrL::gfp,* and *iciA::gfp* (Fig. 4.5C, upper row), a slight regulation for *ybdH::gfp, serA::gfp, ilvE::gfp* and *ndk::gfp,* but almost no regulation for *brnQ*. However, an $\approx$3-fold increase of chromosomal *ndk* mRNA level was observed on Northern blots for a *Salmonella gcvB* deletion strain at mid-log growth compared to the isogenic wild-type strain (data not shown), giving an additional evidence for an influence of GcvB RNA. The proposed interactions as well as locations of C/A-rich elements are shown in Figure 4.4D. Of these new targets, *brnQ* encodes for an amino acid transporter, *ilvC, ilvE* and *serA* for genes involved in amino acid metabolism or biosynthesis, and *thrL* for the *thr* operon leader peptide.

4.1.5. GcvB represses expression of the glycine transporter CycA

One target, *cycA,* was the only one which is repressed in all microarray sets (Table 4.1, indicated in green). CycA belongs to the superfamiliy of APC (amino acid/ polyamine/ organocation) transporters and acts as a permease for glycine, D-alanine, D-serine, and D-cycloserine in *E. coli* (Cosloy, 1973; Robbins & Oxender, 1973; Russell, 1972; Wargel *et al.*, 1971). In *Salmonella* wild-type + C *vs* Δ*gcvB* + C, *cycA* mRNA is repressed $\approx$3.4-fold. Upon pulse-overexpression of GcvB wild-type, $GcvB_{\Delta R1}$, and $GcvB_{\Delta R2}$ mutant, *cycA* mRNA levels drop to $\approx$20%, 28%, and 17%, respectively. Thus, although a slight reduction of regulation was observed for the $GcvB_{\Delta R1}$ mutant, regulation of this target seems to be independent of consensus R1 and R2 indicating alternative binding sites for GcvB RNA. This pattern of regulation was confirmed for a *cycA::gfp* fusion (Fig. 4.6A and B). Quantification of regulation on Western blots also indicated a slight loss of regulation for the $GcvB_{\Delta R1}$ mutant compared to GcvB wild-type and the R2 deletion (Fig. 4.6B). This effect was not visible during FACS analysis. Furthermore, also a double mutant $GcvB_{\Delta R1\,\&\,\Delta R2}$,

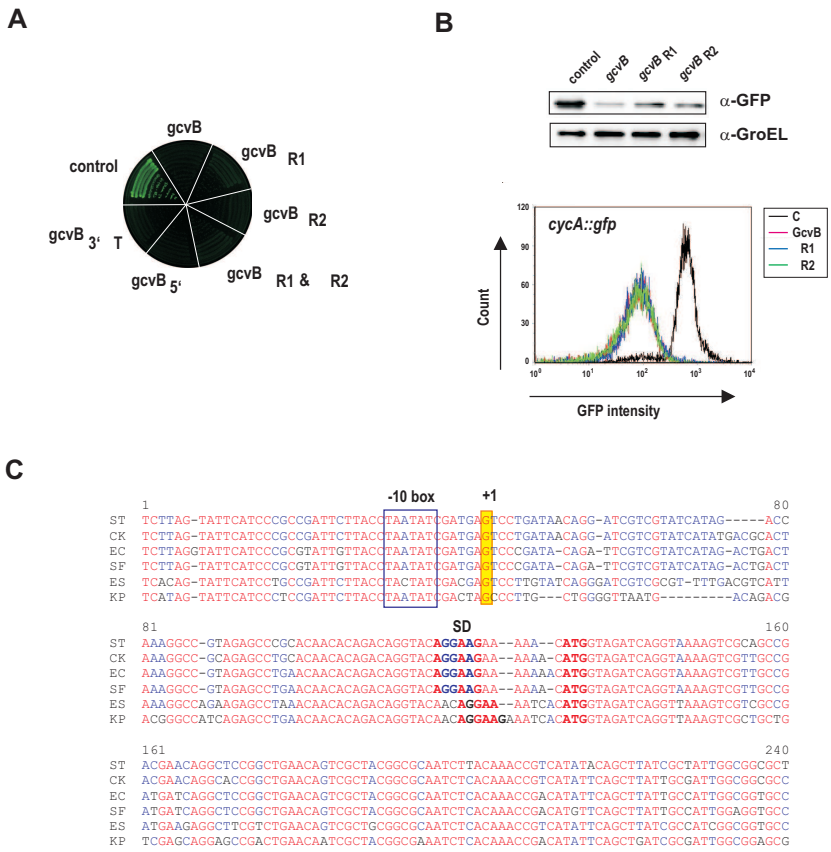


Figure 4.6: Regulation of the glycine/D-alanine/D-serine permease CycA by GcvB. (A) Agar plate-based assay of colony fluorescence of *E. coli* Δ*gcvB recA*$^-$ strains carrying the *cycA::gfp* (pJL30-14) fusion in combination with control vector pTP11, or plasmids expressing *Salmonella* wild-type GcvB RNA (pP_L*gcvB*) or five of the mutant alleles (plasmids pP_L*gcvB*$_{\Delta R1}$ or pP_L*gcvB*$_{\Delta R2}$, pP_L*gcvB*$_{\Delta R1\Delta R2}$, pP_L*gcvB*$_{5'\Delta}$, or p*gcvB*$_{3'\Delta T}$). **(B)** Downregulation of *cycA::gfp* by GcvB wild-type and mutant RNAs GcvB$_{\Delta R1}$ and GcvB$_{\Delta R2}$ after growth to stationary phase was confirmed on Western blots and by FACS analysis. Fold changes of GFP fusion protein levels on Western Blots (upon normalization to GroEL levels) by *gcvB*, *gcvB*$_{\Delta R1}$, or *gcvB*$_{\Delta R2}$ co-expression relative to the control plasmid were: -8.3/-3.4/-6.3. The different colours in the FACS diagram indicate plasmids expressing *Salmonella* wild-type GcvB RNA (pP_L*gcvB*, red), two of the mutant alleles (pP_L*gcvB*$_{\Delta R1}$, blue; pP_L*gcvB*$_{\Delta R2}$, green), or a control plasmid (black). **(C)** Alignment of promoter regions and N-terminal coding sequences of *cycA* homologues in diverse enterobacteria. The transcriptional start site mapped by 5' RACE is indicated by a yellow box and the putative -10 box framed in blue, respectively.

as well as 5' end and 3' end truncated GcvB mutants, GcvB$_{5'\Delta}$ and GcvB$_{3'\Delta T}$, which lack residues 1-91 (SL1 and R1) or 135-201 (R2 to SL5), are still able to repress translation of the *cycA::gfp* fusion (Fig. 4.6A). Furthermore, quantitative real-time PCR analysis of RNA harvested at an OD_{600} of 0.4 indicated an $\approx$3.5-fold upregulation of *cycA* mRNA in a *gcvB* deletion strain compared to the *Salmonella* wild-type strain (data not shown). This is in agreement with the 3.4-fold change between the *Salmonella* wild-type and *gcvB* deletion strain observed in the microarray (Table 4.1, column WT + C).

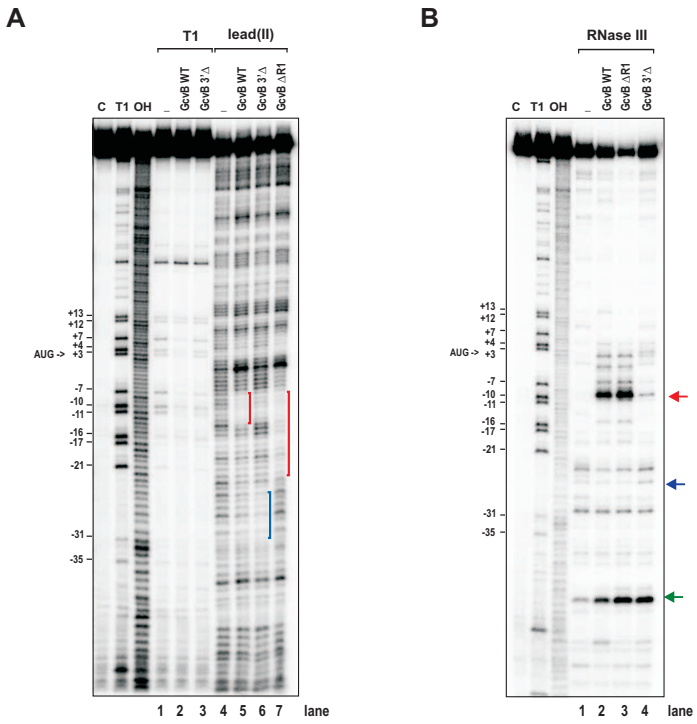


Figure 4.7: Identification of GcvB binding sites on *cycA* mRNA by *in vitro* probing. (A, B) 5' end-labelled *cycA* RNA (≈ 5 nM) treated with RNase T1, lead(II) **(A)**, or RNase III **(B)**. The synthesized *cycA* RNA fragment comprises region -163/+72 relative to the AUG start codon. The 'G' of the AUG start codon corresponds to position '+3'. (Lane C) Untreated *cycA* RNA. (Lane T1) RNase T1 ladder of hydrolysed denatured *cycA* RNA. The position of cleaved G residues is given left of the gel. (Lane OH) Alkaline ladder. **(A)** Probing of *cycA* in the absence (lane 1 and 4) or presence of GcvB wild-type (lanes 2 and 5), $GcvB_{\Delta R1}$ (lane 7), or $GcvB_{3'\Delta}$ mutant RNAs (lanes 3 and 6); final concentrations in lanes 2, 3, 5, 6, 7: ≈ 500nM . Protected regions in *cycA* mRNA are indicated by red and blue bars and RNase III cleavage sites **(B)** in the same regions by red and blue arrows, respectively. The green arrows denote an additional specific RNase III cleavage site of *cycA* RNA in the presence of GcvB wild-type or mutant RNAs.

4.1.6. GcvB binds to *cycA* mRNA *in vitro*

To map the GcvB interaction sites, the RNA of the previously cloned *cycA* fragment was synthesized *in vitro* and subjected to structure probing experiments. RNA structure probing with RNase T1 and lead(II) acetate showed that the presence of GcvB wild-type or mutant RNAs results in 'footprints' on 5'-end-labelled *cycA* leader (Fig. 4.7A). GcvB wild-type protects around seven residues and $GcvB_{\Delta R1}$ ≈ 18 residues starting at position -7 according to the start codon of *cycA* (Fig. 4.7A, red bars). Deletion of the 3' end of GcvB RNA in the $GcvB_{3'\Delta}$ mutant RNA leads to an additional protected region of around eight nucleotides more upstream in the 5' UTR (Fig. 4.7A, blue bar). These interactions were supported by specific RNase III cleavages for each of the GcvB forms (Fig. 4.7B).

The reciprocal experiment, *i. e.*, probing of labelled GcvB RNA in the presence of *cycA* leader, identified multiple *cycA* binding-sites on GcvB RNA (Fig. 4.8A) which were supported by strong

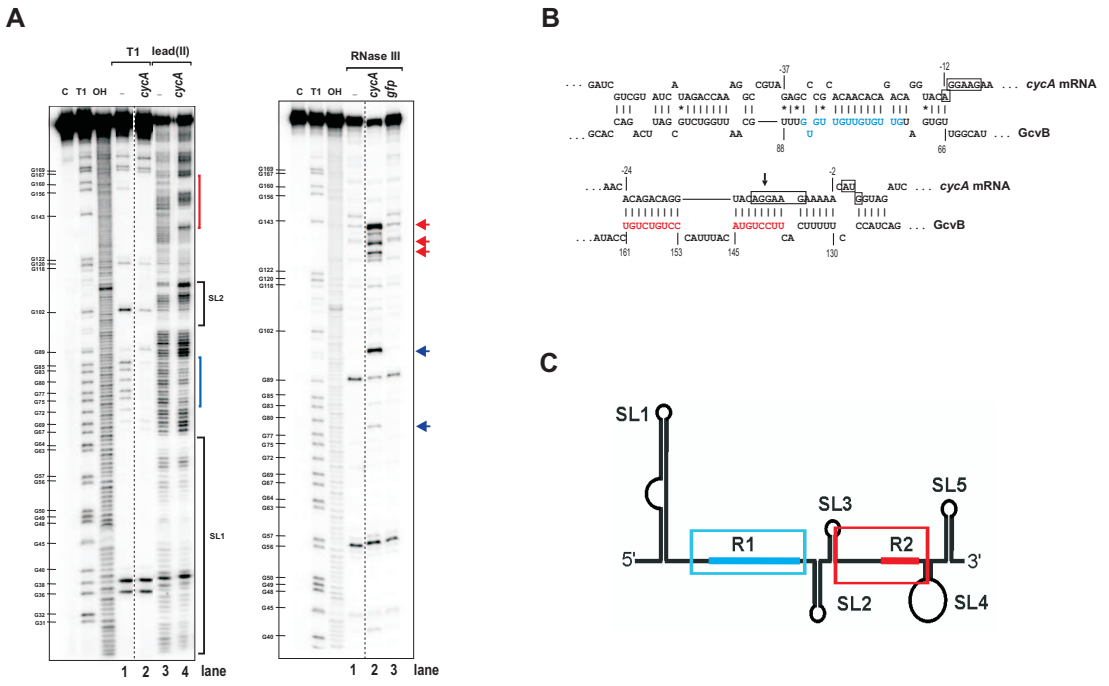


Figure 4.8: Identification of *cycA* binding sites on GcvB RNA by *in vitro* probing. (A) 5' end-labelled GcvB RNA ($\approx$5 nM) was subjected to RNase T1, lead(II) in the absence (lanes 1 and 3) or presence of cold *cycA* RNAs (final concentration in lanes 2 and 4: $\approx$500 nM) or duplexes were confirmed by RNaseIII cleavages in the absence (lanes 1) or presence of cold *cycA* or as a negative control, *gfp* mRNA leader (final concentration in lanes 2 and 3: $\approx 1\ \mu$M). (Lane C) Untreated GcvB RNA. (Lane T1) RNase T1 ladder of hydrolyzed denatured GcvB RNA. The position of cleaved G residues is given left of the gel. (Lane OH) Alkaline ladder. Protected regions are indicated by red and blue vertical bars and RNase III cleavage sites by red and blue arrows, respectively. The approximate positions of stem-loop structures SL1 and SL2 according to the GcvB RNA structure shown in **C** are indicated to the right of the gel. Note that parts of the original gel with unrelated samples were cut out of the picture (indicated by a dashed line). **(B)** Two proposed interaction sites of GcvB-*cycA* complexes. SD and AUG start codon sequences are boxed. The coloured residues were protected from lead(II) cleavage upon duplex formation (see **A**). **(C)** Secondary structure model of GcvB RNA and location of the G/U-rich consensus R1 (blue bar) and consensus R2 (red bar). Two interaction sites derived from the structure probing experiments in **(A)** and predictions in **(B)** are indicated by red and blue boxes, respectively.

RNase III cleavages (Fig. 4.8B). At least two binding sites could be derived from the structure probing experiments and predictions for interactions: one involving consensus R1 (Fig. 4.8B and C, marked in blue) and the other involving R2 (Fig. 4.8B and C, marked in red). The first interaction involves a C/A-rich element upstream of the SD sequence in the *cycA* 5' UTR, the second ovelaps the SD and start codon of *cycA*.

4.1.7. Diverse GcvB mutants indicate multiple binding sites for *cycA* mRNA

To define a minimal GcvB fragment which represses the *cycA::gfp* fusion, diverse GcvB mutant alleles were constructed by subsequent shortening (as described below) of GcvB fragments that are

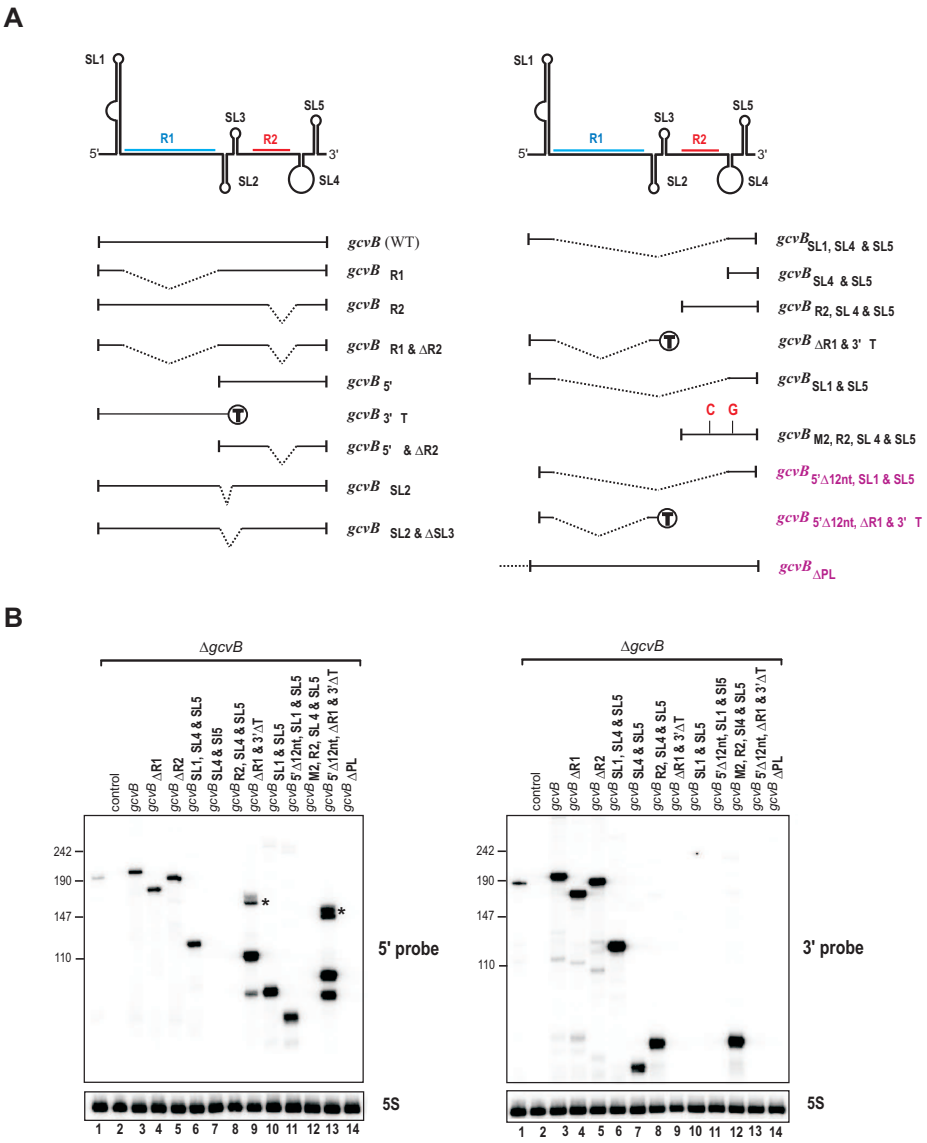


Figure 4.9: Diverse mutant alleles of *gcvB* and their expression. (A) Horizontal bars below two schematic drawings of GcvB RNA denote the *gcvB* fragments expressed by mutant alleles; dotted lines denote internal deletions. Details of the parts that are deleted for each mutant are given in the Chapter Material and Methods in Tables 8.11 and 8.12. In allele $gcvB_{3'\Delta T}$, which derives from a $gcvB_{3'\Delta}$ mutant, SL3 was modified to a transcription terminator. Furthermore, all mutant alleles that carry the $gcvB_{3'\Delta T}$ mutant part are expressed from the native *gcvB* promoter. The mutant allele $gcvB_{\Delta PL}$ lacks the constitutive P_{LlacO} promoter region. Mutant alleles that are indicated in magenta do not repress expression of the *cycA::gfp* fusion. **(B)** Northern blots showing expression of *gcvB* or mutant alleles from a mid-copy plasmid under control of the P_{LlacO} promoter (or *gcvB* promoter for 3' end truncations) in *E. coli* Δ*gcvB* (lanes 3-13). Deletion of the P_{LlacO} promoter region leads to a loss of GcvB expression (lane 14). RNA was isolated from *E. coli* grown to an OD_{600} of 1, and except lane 1 (TOP 10 wild type; JVS-2000) from a Δ*gcvB* genetic background (JVS-6081). The strains in lanes 2 to 14 carried control vector, pTP11, or GcvB mutant plasmids as indicated. Marker sizes are shown to the left. The left blot was probed with labelled oligo JVO-0749, which is complementary to the GcvB 5' region (20-4 bp); the right blot with JVO-0750 complementary to bp 172-150. The asterisks denote transcriptional read-through to the *rrnB* terminator located downstream of *gcvB* on the plasmids.

still able to regulate the *cycA::gfp* fusion (Fig. 4.9A). Plasmids carrying these alleles under control of the constitutive P_{LlacO} promoter expressed distinct GcvB-derived RNAs at levels similar to wild-type *gcvB* (Fig. 4.9B). To rule out any influence of the DNA region itself encoding GcvB RNA, a plasmid carrying a promoter deletion mutant, p*gcvB*$_{\Delta PL}$, was constructed where no GcvB RNA was expressed. To investigate which of the *gcvB* mutant alleles can repress expression of the *cycA::gfp* fusion, each of the GcvB mutant plasmids was co-transformed with the *cycA::gfp* fusion plasmid in *E. coli* Δ*gcvB* and assayed for colony fluorescence on agar plates (Figure 4.10A). Regulation of *cycA::gfp* expression by the different mutant RNAs was quantified by FACS analysis and confirmed the regulation pattern observed on agar plates (Figure 4.10B).

Starting from the GcvB wild-type sequence, deletion of consensus R1 or R2 alone or in combination, as well as 5' end or 3' end truncations of GcvB RNA had no impact on regulation of the *cycA::gfp* fusion as previously shown in Figure 4.6A (Fig. 4.10A and B, number 3 to 7). In contrast, the promoter deletion (*gcvB*$_{\Delta PL}$) does not repress GFP expression of the *cycA::gfp* fusion and indicates that the DNA sequence of the *gcvB* gene itself has no effect on regulation of *cycA* (Fig. 4.10A and B, number 8).

Next, the impact of the two internal stem-loop structures SL2 and SL3 on GcvB regulation was examined, but their deletion had also no influence on repression of the *cycA::gfp* fusion (Fig. 4.10A and B, 9 and 10). As both proposed interactions in Figure 4.8B cover extended regions including consensus R1 and R2, deletion of R1 and R2 might not be sufficient to abolish *cycA::gfp* regulation, and thus, 5' end and 3' end truncations were combined with R2 or R1 deletions, respectively. However, also these GcvB truncations still regulated *cycA* (Fig. 4.10A and B, 11 and 17, pP_L*gcvB*$_{5'\Delta\,\&\,\Delta R2}$ and p*gcvB*$_{\Delta R1\,\&\,3'\Delta T}$). Even further shortened GcvB mutants, where only SL1, SL4, and SL5 (pP_L*gcvB*$_{SL1,\,SL4\,\&\,SL5}$) or consensus R2, SL4, and SL5 (pP_L*gcvB*$_{R2,\,SL4\,\&\,SL5}$) were present, did not abolish *cycA* regulation (Fig. 4.10A and B, 12 and 13). A plasmid habouring only SL4 and SL5 (pP_L*gcvB*$_{SL4\,\&\,SL5}$) turned out to be instable (data not shown). Although in plasmid pP_L*gcvB*$_{R2,\,SL4\,\&\,SL5}$ already 134 nt of GcvB are deleted, it could still form the second proposed interaction which involves base-pairs between consensus R2 and the *cycA* SD sequence and a second helix that was confirmed by structure probing (Fig. 4.8B, red interaction). To disrupt these interaction sites, a single-nucleotide exchange was introduced in each of these helices (pP_L*gcvB*$_{M2,\,R2,\;\,SL4\,\&\,SL5}$). Although this led to a slight loss of regulation (Fig. 4.10A and B, 14), the GcvB fragment seems to be still able to bind to *cycA* mRNA. Similarily, deletion of residues 66-177 (pP_L*gcvB*$_{SL1\,\&\,SL5}$) leads only to a partial loss of regulation, although both proposed interaction regions are completely deleted (Fig. 4.10A and B, 15). This indicates that GcvB probably carries multiple sites which can interact with *cycA* mRNA.

However, further deletion of the first 12 nt from the 5' end in allele *gcvB*$_{\Delta 5'12nt,SL\,1\,\&\,5}$ leads to a complete loss of regulation (Fig. 4.10A and B, 16). Similarily, deletion of these 12 first nucleotides from the p*gcvB*$_{\Delta R1\,\&\,3'\Delta T}$ plasmid resulting in allele p*gcvB*$_{5'\Delta 12nt,\,\Delta R1\,\&\,3'\Delta T}$, abolished more than half of the *cycA* regulation (Fig. 4.10A and B, 18). The first 12 nt from the 5'-end of GcvB RNA are predicted to bind to the SD sequence of *cycA* (Fig. 4.10B). Single base-pair mutations at position 3, 8, or 11 introducted into the *gcvB*$_{SL\,1\,\&\,5}$ allele are sufficient to abrogate regulation by the GcvB mutant RNA (Fig. 4.10D). Overall, this indicates at least three major GcvB interaction sites for binding of *cycA* mRNA around the SD sequence.

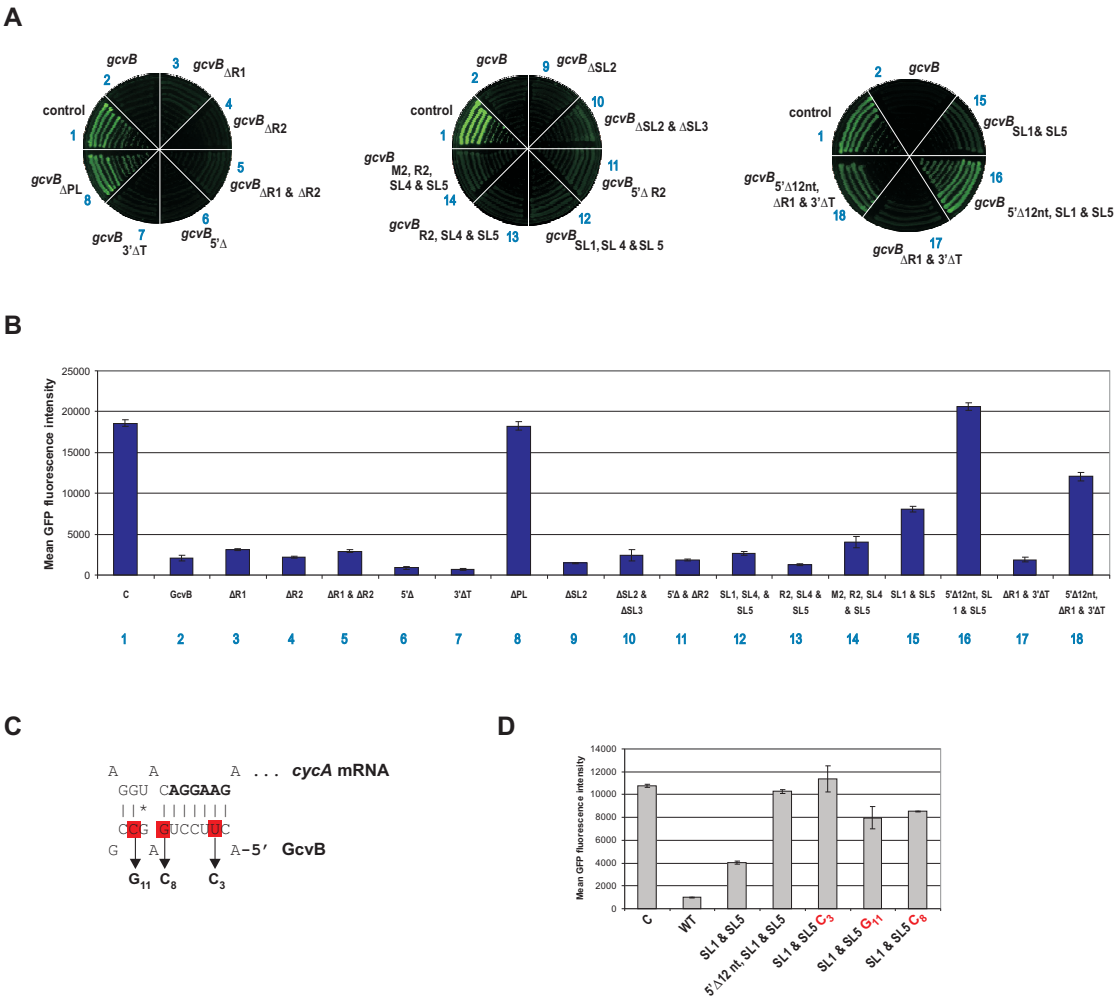


Figure 4.10: Diverse GcvB mutants can regulate *cycA*. (A) Agar plate-based assay of colony fluorescence of *E. coli* Δ*gcvB recA*$^-$ strains carrying the *cycA::gfp* (pJL30-14) fusion in combination with control vector pTP11 (1), or plasmids expressing *Salmonella* wild-type GcvB RNA (pP_L*gcvB*) (2), a promoter deletion mutant *gcvB*$_{\Delta PL}$ (8), or 15 of the *gcvB* mutant alleles (2-7, and 9-18). **(B)** Regulation of *cycA::gfp* expression by the different mutant RNAs was quantified by FACS analysis. *E. coli* Δ*gcvB recA*$^-$ strains carrying the *cycA::gfp* reporter in combination with the sRNA plasmids were grown for 14 hours and analyzed by FACS. Reporter activity is given in arbitrary units (relative GFP fluoresence) and plotted as average from two independent experiments including error bars. Blue numbers indicate sRNA plasmids as in **(A)**. **(C)** Proposed interaction between the first 12 nucleotides of GcvB RNA and the *cycA* SD sequence (indicated by bold letters). Single base-pair exchanges that were introduced in GcvB at position 3, 8 and 11 are indicated by arrows. **(D)** *E. coli* Δ*gcvB recA*$^-$ strains carrying the *cycA::gfp* reporter in combination with the control plasmid pTP011, or plasmids carrying the *gcvB* wild-type or mutant alleles, *gcvB*$_{SL\ 1\ \&\ 5}$, *gcvB*$_{\Delta 5'12nt,SL1\ \&\ 5}$, and single base-pair exchanges in *gcvB*$_{SL1\ \&\ 5}$, were grown for 14 hours and analyzed by FACS. Reporter activity is given in arbitrary units (relative GFP fluoresence) and plotted as average from two independent experiments including error bars.

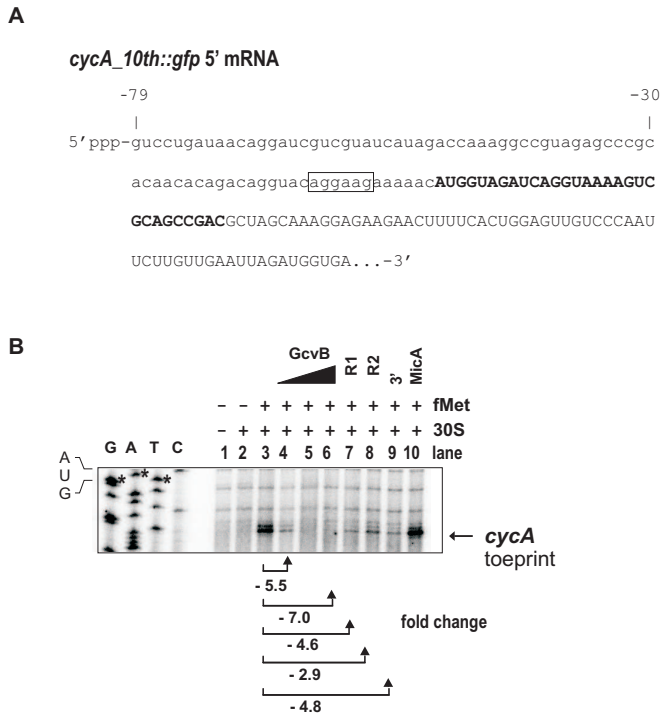


Figure 4.11: GcvB inhibits 30S binding to *cycA* mRNA *in vitro*. (A) An *in vitro* synthesized *cycA::gfp* fusion mRNA fragment was used in the toeprint assay. The SD sequence is indicated by a box. Uppercase letters indicate the coding sequence, whereof bold letters correspond to the *cycA* fusion part and plain letters to the *gfp* fusion part, respectively. **(B)** Ribosome toeprinting of *cycA::gfp* leader RNA (20 nM) as described in Material and Methods. '+/-' indicate the presence or absence of 30S subunit (200 nM) and fMet initiator tRNA (1 μM). The *cycA::gfp* AUG start codon position is shown. The arrow indicates the 30S toeprint. Increasing concentrations of GcvB RNA (lanes 4-6: 20, 100 and 200 nM) in the reactions inhibit 30S binding whereas the unspecific control RNA, MicA (lane 10, 200 nM) does not impair binding. Mutant RNAs $GcvB_{\Delta R1}$, $GcvB_{\Delta R2}$, and $GcvB_{3'\Delta}$ (lanes 7-9) were added at a final concentration of 200 nM. Fold-repression values for the different GcvB forms and concentrations are given below the gel.

4.1.8. GcvB inhibits translation initiation of *cycA* mRNA *in vitro*

GcvB RNA can bind with multiple sites near the Shine-Dalgarno sequence of *cycA* (Fig. 4.8B and 4.10C), and was thus predicted to prevent ribosome binding to this mRNA. To test this, 30S ribosome toeprinting assays (Hartz *et al.*, 1988) were performed. A *cycA_10th::gfp* fusion mRNA fragment (10^{th} amino acid of *cycA* was fused to *gfp*, Fig. 4.11A) was annealed to an end-labelled primer complementary to the *cycA::gfp* coding region (+68 to +92), and incubated with 30S subunits in the presence or absence of uncharged $tRNA^{fMet}$, followed by cDNA synthesis. Analysis of the extension products (Fig. 4.11B) revealed one ribosome-induced, $tRNA^{fMet}$-dependent termination site at the characteristic +15/+16 positions (start codon A is +1). This 'toeprint' signal was decreased when increasing concentrations of wild-type GcvB RNA were added prior to incubation with 30S/fMet (lanes 4-6), suggesting inhibition of 30S binding. In contrast, the unrelated MicA RNA (lane 10) did not inhibit ternary complex formation.

Inhibition was also observed with the GcvB$_{\Delta R1}$, GcvB$_{\Delta R2}$ and GcvB$_{3'\Delta}$ mutant RNAs, but to a lesser extent (lanes 7-9). Specifically, a 10-fold excess of GcvB wild-type reduced 30S binding to the *cycA_10th::gfp* fusion mRNA $\approx$7-fold, whereas the mutant RNAs GcvB$_{\Delta R1}$, GcvB$_{\Delta R2}$, and GcvB$_{3'\Delta}$ reduced 30S binding only 4.6-, 2.9-, and 4.8-fold, respectively. Although this indicates different binding affinities of the different GcvB variants to *cycA* mRNA, probably all of them act as inhibitors of translation initiation by masking the *cycA* RBS.

4.2. Discussion

It has become increasingly clear that bacterial sRNAs regulate multiple rather than individual mRNAs. Early studies often focussed on the regulation of a single target for a given sRNA; in fact, *ompF* mRNA has remained the only investigated target of MicF sRNA in 25 years (Mizuno *et al.*, 1984). However, several studies in *E. coli* and *Salmonella* revealed extended regulatory networks that rely on multiple-targeting by bacterial sRNAs (see Section 2.5). For example, the OmrA/B and RybB were reported to target multiple mRNAs that collectively encode for outer membrane proteins (Guillier & Gottesman, 2006; Johansen *et al.*, 2006; Papenfort *et al.*, 2006) or the iron stress-responding *E. coli* RyhB sRNA was shown to regulate multiple mRNAs that encode proteins involved in iron metabolism (Massé & Gottesman, 2002). Though, due to the low number of validated sRNA-target interactions the mechanistic details of how sRNAs could directly control multiple mRNAs by antisense mechanism have remained largly unclear. In addition, reliable prediction of new targets is still a challenging task.

In the previous Chapter it was shown that GcvB RNA from *Salmonella* directly interacts by a highly conserved G/U rich region with C/A-rich elements in the 5' UTRs of seven ABC-transporter mRNAs and thereby represses translation. The approaches for the identification of GcvB targets in the previous Chapter were restricted to periplasmic proteins, as *E. coli* GcvB RNA has been shown to regulate several periplasmic transporters (Urbanowski *et al.*, 2000). Here, an extended search for additional GcvB targets by global analysis of mRNA changes on microarrays after GcvB pulse-overexpression was undertaken to analyse the GcvB regulon. Pulse-overexpression of sRNAs has previously been used for the successful identification mRNA targets (Section 2.4.1.2). In addition to wild-type GcvB RNA, also global mRNA changes upon pulse-expression mutant RNAs lacking the consensus sequences R1 and R2 were analysed. In these GcvB microrray experiments, almost half of the known GcvB targets from the previous Chapter (*dppA*, *oppA*, and *gltI*) were downregulated by GcvB along with the downstream genes of the corresponding operons (Table 4.1). Furthermore, the previously characterized R1-dependent regulation of these targets was also visible in the microarray experiments as regulation of these targets was abolished upon pulse-expression of the GcvB$_{\Delta R1}$ mutant (Table 4.1, compare columns 'Δ*gcvB* + GcvB' and 'Δ*gcvB* + ΔR1'). In contrast, pulse-expression of the GcvB$_{\Delta R2}$ mutant RNA leads to comparable mRNA changes as for GcvB wild-type (Table 4.1, compare columns 'Δ*gcvB* + GcvB' and 'Δ*gcvB* + ΔR2'), except for *gltI* mRNA which is $\approx$2-fold more downregulated upon pulse-expression of the GcvB$_{\Delta R2}$ mutant RNA. Deletion of R2 could change the affinity of GcvB for several other targets by removing an additional binding site and increase the affinity for this target.

Pulse-expression of the sRNA should avoid the pleiotropic effects that can be expected to result from constitutive sRNA expression and allows identification of direct targets, *i. e.* mRNAs which are

directly bound by a given regulatory RNA due to short expression times. For example, chromosomal GcvB itself showed an ≈ 2-fold increase after arabinose addition, however, induction of GcvB RNA from the plasmid led to ≈ 2.5-fold higher level than the chromosomal GcvB after 10 min. Thus the pulse-expression is probably comparable to wild-type *gcvB* expression. This is also reflected by ≈ 7-fold changes for *dppA* and *oppA* mRNAs by chromsomal GcvB RNA compared to ≈ 13-fold and ≈ 18-fold changes by GcvB wild-type expressed from the pBAD-plasmid (Table 4.1).

In addition, similar levels of regulation by chromosomal GcvB RNA were observed in the microarray experiment (Table 4.1, *dppA*: -6.8-fold, *oppA*: -7.3-fold, and *gltI*: -4.2-fold) compared to the fold-changes determined by quantitative real-time (qRT) PCR in Figure 3.13C in Section 3.1.4 (*dppA*: ≈ 12-fold, *oppA*: ≈ 8-fold, and *gltI*: ≈ 6-fold). In contrast, *livJ* and STM4351 which showed ≈ 6-fold and 4.5-fold downregulation by GcvB in the qRT-PCR experiment (Figure 3.13C) are not or only slightly (-1.6-fold for STM4351) downregulated in the microarray experiment. This might be due to different or low expression of these genes in the growth conditions used for these experiments (OD_{600}=1 in the microarray and OD_{600}=0.4 for qRT PCR, respectively). Both remaining known targets, *argT* and *livK*, did also not reach the 2-fold threshold in all microarray experiments. However, as these genes showed also only slight mRNA changes in the qRT PCR in Figure 3.13C, GcvB probably mainly inhibits translation of these genes without promoting RNA degradation.

Including also the GcvB mutant RNAs in the pulse-expression approach, allowed to identify further consensus R1-dependent targets. All R1-dependent targets from the previous Chapter are bound at C/A-rich regions by GcvB RNA and also the new targets from the microarray approach were predicted to have C/A-rich target sites. Thus this extendend set of R1-dependent targets was used to define a consensus motif for the GcvB R1 target site using `MEME`. Motif-searches for the presence of the consensus motif in the 5' regions of all *Salmonella* RNAs in combination with predictions for potential interactions using `RNAhybrid` (Rehmsmeier *et al.*, 2004) clearly improved target predictions and revealed additional R1-dependent GcvB targets. In total, >30 potential GcvB consensus R1-dependent targets were identified in the two approaches, whereof at least ten could be validated by Western blots or FACS analysis of GFP reporter-gene fusions (*cycA, ygjU(sstT), gdhA, yaeC, ilvC, thrL, iciA, ybdH, serA*, and *brnQ*,). Recently, additional GcvB targets were also described in *E. coli* (Pulvermacher *et al.*, 2009a,b) based on a microarray analysis of RNA isolated from wild-type and a *gcvB* deletion strain in a comparable growth condition as in this study. A large overlap of regulated genes was found between *E. coli* and *Salmonella* indicating a conserved regulatory function of GcvB RNA. Specifically, the L-serine and L-threonine transporter *sstT* (*ygjU*) is also repressed in *E. coli* (Pulvermacher *et al.*, 2009b), but also *gdhA, serA, ilvB*, and *ilvC* were up-regulated besides the previously known targets in the *E. coli gcvB* deletion strain (Pulvermacher *et al.*, 2009a).

For some of the targets, *e. g.*, *asd* derived from the microarray data (Fig. 4.3), regulation could not be confirmed by GFP-fusions although the fusion contained a region which was predicted to interact with GcvB (Fig. 4.4A). Thus, either it is not a direct target or the cloned region is not sufficient for GcvB binding. For example, OxyS RNA binds *fhlA* mRNA at two interaction sites (Altuvia *et al.*, 1998). The 5' parts of the putative GcvB targets that were cloned in the GFP-fusions started at transcriptional start sites identified by 5'RACE or at promoters derived from the literature. However, alternative transcriptional start sites could lead to longer/ shorter transcripts which are targeted by GcvB. Transcripts of different length could also fold into different secondary structures which could

be necessary for efficient GcvB binding. For some GFP fusions, regulation was observed by monitoring colony-fluorescence on agar plates but not in Western blot or FACS experiments. One reason for these discrepancies in regulation could be the different growth conditions between agar plates and liquid cultures, which were used for Western blot and FACS analysis. In addition, expression of an additional trans-acting factor, such as Hfq, could be required for regulation. Furthermore, the constitutive expression of the sRNA could lead to pleiotropic effects, such as Hfq titration which was previously reported for OxyS RNA (Zhang *et al.*, 1998) and is also proposed for SraH RNA in *Salmonella* (K. Papenfort and J. Vogel, unpublished). This could unintentionally change the stability of other sRNAs that are expressed under a specific growth condition which in turn affects the expression of their target mRNAs. In addition, the expression of larger sets of genes can be altered by targeting the mRNAs of global transcriptional regulators, *e. g.* RpoS or FhlA (Argaman & Altuvia, 2000; Lease *et al.*, 1998; Majdalani *et al.*, 1998).

Regulation of one target that was identified in the microarray experiment, *cycA* mRNA, turned out to be independent of the G/U-rich consensus R1, which is strictly required for regulation of all other identified targets. Furthermore, it turned out that diverse GcvB mutants are still able to repress *cycA* mRNA, indicating multiple, alternative binding sites within GcvB RNA for this target. Toeprinting analysis confirmed that GcvB RNA can directly inhibit translation initiation of *cycA* mRNA. In addition, *cycA* was also identified in the above mentioned microarray-based approach as a GcvB target in *E. coli* (Pulvermacher *et al.*, 2009a). In the same study it turned out that diverse mutations within GcvB RNA have no impact on *cycA* regulation. Here, at least three independent binding sites could be identified within GcvB RNA but it is not clear which of the different binding sites actually bind the *cycA* mRNA *in vivo*. The fact that diverse GcvB mutants are still able to repress the *cycA::gfp* fusion mRNA to the same level could be an effect of the overexpression from the plasmid. Whereas regulation of the *cycA::gfp* fusion mRNA seems to be comparable for, *e. g.*, GcvB wild-type and the R1 and R2 deletion mutants, toeprinting analysis indicated a reduced inhibition of translation initiation by the mutant RNAs. Further toeprinting experiments, with several of the shorter GcvB mutants could give additional hints at the amount of regulation by the different fragments. Furthermore, integration of the different mutants into the *Salmonella* chromosome under control of the native *gcvB* promoter could help to avoid these overexpression effects.

The *cycA* gene is particularly interesting as a GcvB target because of its function as a glycine transporter. Stauffer & Stauffer (2005) showed that the transcriptional activator GcvA binds to a single region in the *gcvB* promoter and binding of GcvA to this region is required for both GcvA-mediated activation in the presence of glycine and GcvA-GcvR mediated repression in the absence of glycine in the medium. Regulation of *cycA* might be important as it is a glycine transporter, and thus regulation might provide a negative feedback/autoregulatory loop for GcvB expression which will be the subject of further investigations.

Besides several downregulated targets, a large fraction of genes was upregulated in the different microarray sets. Especially the *pyr* genes showed strong elevations in mRNA levels (up to $\approx$ 90-fold) upon GcvB pulse-expression (Table 4.1). However in qRT PCR experiments, the amount of, *e. g.*, *pyrB* mRNA, was not significantly changed in a *gcvB* deletion strain compared to the *Salmonella* wild-type strain indicating these genes as false-positives (data not shown). Furthermore, these genes and also, *e. g.*, the *fim* genes are known to be highly sensitive to small changes in growth conditions

as they show aberrant changes in diverse microarray experiments (K. Papenfort and S. Lucchini, personal communication). Thus, these genes have a higher probability of being false positives and were therefore excluded from further analysis. Nevertheless, it could be intersting to test if several of the other activated genes are directly upregulated by GcvB RNA as sRNA-mediated activation seems to be more widespread than previously anticipated. Besides the first example, DsrA, several other sRNAs have been shown to activate gene expression, *e. g.*, by disrupting an inhibitory secondary structure which masks the ribosome binding site and thereby prevents translation in the absense of the sRNA (see Section 2.1.2). Also in *E. coli*, diverse genes were upregulated by GcvB RNA, but none of them overlapped with the genes identified in this study (Pulvermacher *et al.*, 2009a).

Pulse-expression of the GcvB consensus R2 deletion mutant changed the expression of $>$150 mRNAs which were not regulated in the three other microarray sets (Fig. 4.2). Although further evidence was obtained for regulation of some of these targets (*e. g. ndk* and *serA*), most of these targets are probably due to 'off-target' effects. The effect of 'off-targeting' has also been observed for siRNAs as a sequence-specific effect by binding unintended transcripts in eukaryotes. This unspecific binding cannot be reliably predicted nor avoided as there is relatively high tolerance for mismatches between a short RNA and its target, leading to undesirable regulation of further targets (Svoboda, 2007). This could also be the case for the $GcvB_{\Delta R2}$ mutant, where the pulse-expression could result in unspecific binding of many mRNAs by the G/U-rich region R1 which allows several wobble-base-pairs in addition to canonical Watson-Crick pairing. Thus, consensus R2 might have a function in discrimination of actual GcvB targets from 'off-targets', whereas most of the target affinity is determined by consensus R1. This implies conserved domains with different functions in GcvB RNA. Furthermore, only one R2-dependent operon (STM4535 to STM4540, encoding for a putative PTS permease system and a putative glucosamine-fructose-6-phosphate aminotransferase) that is downregulated by GcvB RNA was detected in the microarrays (Table 4.1). Whether these are really direct targets of GcvB RNA and if R2 is strictly required for binding remains to be further investigated. In eukaryotes, it was shown that transcriptional repression by Alu RNA during the cellular heat shock response involves two loosely structured domains that are modular (Mariner *et al.*, 2008). Also RNAIII from *S. aureus* is a bifunctional molecule that encodes the δ-hemolysin protein in its 5' end while it also acts as a noncoding regulatory RNA, primarily with its 3' domain (Boisset *et al.*, 2007; Huntzinger *et al.*, 2005; Janzon *et al.*, 1989). Similarly, SgrS RNA, which is expressed in *E. coli* during glucose-phosphate stress, encodes a short ORF within its 5' end, whereas the 3' end mediates target interaction with *ptsG* mRNA (Kawamoto *et al.*, 2006; Wadler & Vanderpool, 2007). Thus, a modular architecture consisting of multiple functional domains, a property reminiscent of classical protein transcriptional regulators, can also be found in RNA regulators.

Conserved subregions, or 'domains', in sRNAs which harbour critical residues for multiple interactions seem to be more widespread than previously anticipated. These conserved domains can be an internal region such as the highly conserved G/U rich region of GcvB RNA which directly interacts with C/A-rich elements in the 5' UTRs of seven ABC-transporter mRNAs from *Salmonella*. Also, CyaR RNA uses a conserved internal region with an almost perfect anti-SD sequence for binding of multiple target mRNAs (De Lay & Gottesman, 2009; Johansen *et al.*, 2008; Papenfort *et al.*, 2008). In contrast, other bacterial sRNAs use a conserved 5' end to interact with multiple targets. For OmrA and OmrB, the first nine nucleotides from the 5' end, which are conserved be-

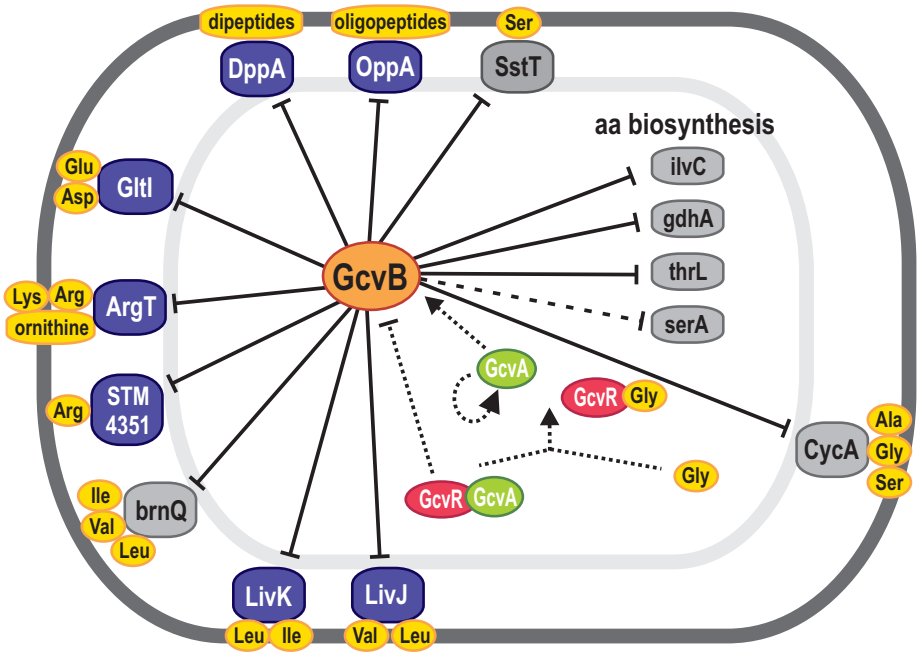


Figure 4.12: GcvB network. GcvB sRNA represses mRNAs of the periplasmic substrate proteins of many ABC transporters involved in amino acid uptake. Seven targets which have been validated by *in vitro* structure probing in the previous Chapter 3 are shown as blue ovals; grey ovals denote additional proteins of transport systems or with roles in amino acid biosynthesis, and whose synthesis has been shown to be directly regulated by GcvB in *Salmonella*. The preferred amino acid, di- or oligopeptide substrates of relevant periplasmic binding proteins are shown in yellow ovals. Control of the *gcvB* gene by two transcription factors (GcvR, GcvA) of the glycine cleavage system is indicated based on extensive work in *E. coli* by the Stauffer laboratory (see, *e. g.*, Urbanowski *et al.*, 2000).

yond *E. coli* and *Salmonella*, were shown to directly recognize multiple target mRNAs (Guillier & Gottesman, 2008). Also RyhB RNA uses 16 nt from its conserved 5' end for direct binding to multiple OMP mRNAs (Bouvier *et al.*, 2008; Mika *et al.*, 2009, *submitted*). The identification of conserved sRNA parts and limiting the sequence space for sRNA-target predictions to these regions will clearly improve bioinformatics-based target identification approaches. However, there are now several mRNAs that are repressed at non-RBS positions, either in the 5' UTR or CDS, and by a diversity of recently deciphered mechanisms (Bouvier *et al.*, 2008; Darfeuille *et al.*, 2007; Heidrich *et al.*, 2007; Vecerek *et al.*, 2007). This leads to larger sequence space for predictions of interactions within the mRNA part. Therefore, additional features, *e. g.*, whether or not an mRNA is bound by Hfq (Sittka *et al.*, 2008; Zhang *et al.*, 2003) or the presence of single-stranded regions that could be accessible for sRNA-target interactions (Busch *et al.*, 2008; Tafer & Hofacker, 2008) will be needed to improve target prediction algorithms.

One of the newly identified targets, *thrL*, encodes for the leader peptide of the *thr* operon. It will be interesting to see if also the downstream genes *thrABC* are downregulated by GcvB RNA. This would be comparable to the case of *fur* regulation by RyhB RNA (Vecerek *et al.*, 2007). Specifically, the *fur* gene is co-transcribed with an upstream ORF (*uof*) and translation of the *uof* is a requisite

for Fur protein synthesis. RyhB RNA targets the RBS of this upstream located short ORF and constitutes a new paradigm of sRNA action. Similarily, the ThrL leader peptide controls expression of the *thrLABC* operon, which encodes four out of the five enzymes of threonine biosynthesis pathway, by an attenuation mechanism in response to threonine and isoleucine levels in *E. coli* (Lynn *et al.*, 1982). Thus, translational inhibition of *thrL* by GcvB RNA is very likely to affect also the downstream operon.

The global target identification approaches showed that GcvB targets not only periplasmic transporters but also genes that are involved in amino-acid metabolism. As almost all of the previous and newly identified targets constitute amino acid or peptide transporters or genes involved in amino acid biosynthesis, this argues for a global regulatory role of GcvB RNA in amino acid metabolism (see Figure 4.12). For this purpose, GcvB regulates several functionally related mRNAs. This has also been observed for other sRNAs, for example RyhB which regulates multiple genes involved in iron metabolism (Massé *et al.*, 2005). Many of the mRNAs demonstrated to be direct targets of *Staphylococcus aureus* RNAIII encode bacterial virulence factors, and OmrA/B and RybB sRNAs were predicted to directly target multiple mRNAs that collectively encode for outer membrane proteins (Guillier & Gottesman, 2006; Johansen *et al.*, 2006; Papenfort *et al.*, 2006). Tjaden *et al.* (2006) suggested that functional relationship of the proteins encoded by target candidates may add confidence to biocomputational target predictions. Furthermore, conserved sRNA parts like the G/U-rich region R1 will help to narrow down the list of target interactions. Increasing numbers of biochemically mapped interactions will then, in turn, improve target predictions. Overall, multiple mRNA targeting by bacterial sRNA turns out to be more common than previously thought and could allow bacteria to rapidly regulate functional modules in response to environmental changes.

CHAPTER 5

ANALYSIS OF HFQ-BOUND RNAS IN *Salmonella* BY 454 PYROSEQUENCING

Although sRNAs have become an important new mediator of bacterial mRNA regulation, biocomputational identfication of novel sRNAs without any conservation information still remains difficult. Recent advances in high-throughput sequencing (HTPS) technology now allow a thorough analysis of RNA bound to cellular proteins and, therefore, of post-transcriptional regulons including sRNA regulators. In bacteria, the majority of sRNAs basepair with target mRNAs to regulate their translation and/or decay, and these regulatory events commonly require the bacterial Sm-like protein Hfq (see Section 2.2.1). Hfq interacts with both regulatory sRNAs and mRNAs, and much of its post-transcriptional function is caused by the facilitation of the generally short and imperfect antisense interactions of sRNAs and their targets (Section 2.1). Furthermore, it is one of the most abundant RNA-binding proteins in bacteria and almost half of all sequenced Gram-negative and Gram-positive species, and at least one archaeon, encode an Hfq homologue (Section 2.2.1). Hfq has an important physiological role in numerous model bacteria, and the pleiotropic effects of an *hfq* deletion mutation as well as an impact of Hfq on virulence have been observed in diverse bacteria (Section 2.2.1). Also in *Salmonella*, reduced virulence and motility as well as deregulation of more than 70 abundant proteins have been observed (Section 2.6.1.1). This indicates an important role of post-transcriptional regulation in *Salmonella* involving probably also *Salmonella*-specific sRNA regulators.

A first step to elucidate the pleiotropic Hfq effects involves the identification of small regulatory RNAs that are involved in the post-transcriptional regulons. Several systematic screens have been developed to identify small RNAs in bacteria (see Section 2.3). However, the bioinformatics-based approaches often rely on the prediction of orphan transcription signals and primary sequence conservation of putative sRNA candidates within closely related species or on conservation of RNA structure (Section 2.3). For some bacteria, the lack of available genome sequences of related bacteria does not allow a thorough conservation analysis. Furthermore, RNAs with a potential role in virulence are probably species-specific due to adaptation to different hosts. Therefore, *de novo* identification of sRNAs remains a major task.

In *Salmonella*, diverse sRNAs have been identified. Many of the *Salmonella* sRNAs, such as GcvB RNA which was described in the two previous Chapters, were initially identified in *E. coli* (see Section 2.6.1.2). As 25% of the total genetic material has been laterally acquired since *Salmonella* diverged from *E. coli* (Porwollik & McClelland, 2003), these *Salmonella*-specific regions could probably encode new sRNAs or other RNA elements that are absent in *E. coli*. Two bioinformatics-based studies (Padalon-Brauch *et al.*, 2008; Pfeiffer *et al.*, 2007) predicted *Salmonella*-specific

sRNAs, which led to the discovery of the first sRNA from an enterobacterial pathogenicity island, *i. e.* the 80 nt InvR RNA that is expressed from the invasion gene locus, SPI-1 (Pfeiffer *et al.*, 2007). Padalon-Brauch *et al.* (2008) showed expression of 19 island-encoded sRNAs under a large panel of growth conditions reminiscent of the environments encountered by *Salmonella* upon host cell infection. Both studies predicted additional *Salmonella* sRNA candidates that remain to be verified experimentally.

One strategy to find novel sRNAs is to look for the sRNA binding partners of known RNA-binding proteins (see Section 2.3.8). As one of the key players in sRNA-mediated regulation, Hfq, is highly conserved in a wide range of bacteria, this protein is a perfect candidate for co-immunoprecipitation of bound RNAs. In a pioneering global study in *E. coli*, Zhang *et al.* (2003) used co-immunoprecipitation with Hfq-specific antisera and direct detection of the bound RNAs on genomic high-density oligonucleotide microarrays. Although this method proved highly effective for detecting diverse sRNAs and mRNAs in *E. coli*, the requirement of high-density microarrays and specialized antibodies has hampered similar studies in other bacteria. An alternative approach identified individual abundant Hfq-associated RNAs by cDNA cloning or direct sequencing (Antal *et al.*, 2005; Christiansen *et al.*, 2006); however, these methods are not appropriate for large-scale analyses due to technical and cost limitations of the conventional Sanger sequencing.

To overcome these limitations for the global identification of Hfq targets in *Salmonella* in this work high-throughput sequencing of RNA associated with an epitope-tagged Hfq protein was used which allows parallel sequencing of hundreds of thousand cDNAs in one sequencing run. Deep sequencing analysis of $\approx$ 350,000 cDNAs, derived from RNA co-immunoprecipitation (coIP) with epitope-tagged Hfq or a control coIP, recovered Hfq-binding sRNAs with high specificity and identified their boundaries with unprecedented resolution. Additionally, many conserved enterobacterial sRNA genes as well as novel *Salmonella* sRNA genes could be detected. In total, the cDNA analysis more than doubled the number of sRNAs known to be expressed in *Salmonella* to 67. Furthermore, 727 genes mRNAs were identified whose transcripts are Hfq-bound *in vivo*. The complete study including also analysis of Hfq-bound mRNAs and validation of the sequencing data by hybridization of Hfq-coIP cDNA on high density microarrays has been previously published in Sittka *et al.* (2008). Furthermore, it includes a transcriptomic analysis of *Salmonella* wild-type and *hfq* deletion mutant strains on microarrays.

In future, the combination of epitope-tagging and HTPS of immunoprecipitated RNA will allow the characterization of sRNAs and mRNAs in different genetic backgrounds or in bacteria grown under various environmental conditions. This approach overcomes the limited availabilty of high-density microarrays that have constrained expression-based sRNA discovery in microorganisms. Thus, this strategy is ideal for the global analysis of the post-transcriptional regulons of RNA-binding proteins and for sRNA discovery in a wide range of genetically tractable bacteria.

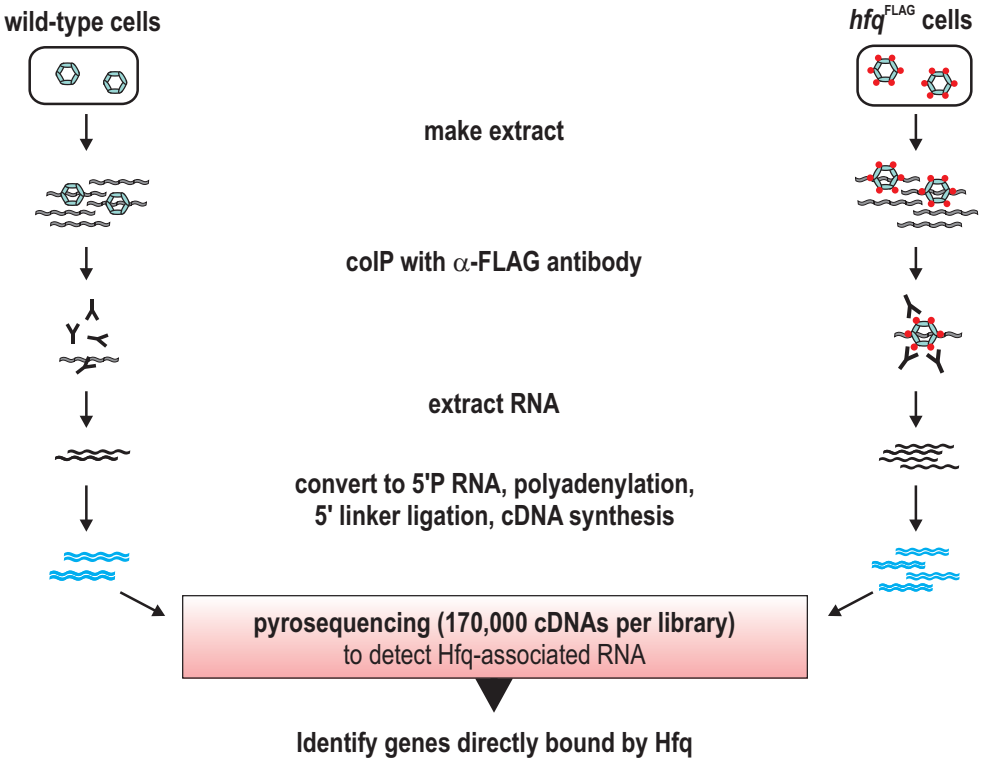


Figure 5.1: Deep sequencing strategy to identify Hfq-bound RNAs. RNA was co-immunoprecipitated with Hfq in extracts from ESP-grown *Salmonella* cells (wild-type and chromosomal *hfq*$^{\mathrm{FLAG}}$ strain) using an anti-FLAG antibody. The extracted RNA was converted to 5' monophosphate RNA, and subsequently into cDNA, followed by direct pyrosequencing. (Figure was adapted from Sittka *et al.*, 2008.)

5.1. Results

5.1.1. Experimental setup for deep sequencing of Hfq-associated RNAs

The pleiotropic effects of an *hfq* mutant indicate the existence of a variety of transcriptional regulons that show Hfq-dependent expression patterns which are either mediated by the binding of certain regulatory sRNAs or of specific mRNAs by Hfq. As a first step to unravel the Hfq regulon, the RNAs that are directly bound by Hfq have to be identified. For this purpose, RNA was co-immunoprecipitated with the chromosomally FLAG epitope-tagged Hfq protein expressed by a *Salmonella hfq*$^{\mathrm{FLAG}}$ strain (Pfeiffer *et al.*, 2007) and the Hfq-bound RNAs were analysed by deep sequencing (see Figure 5.1). Co-immunoprecipitation was performed with an α-FLAG antibody in extracts prepared from early stationary phase (ESP)-grown bacteria of a *Salmonella* wildtype (coIP-Ctr) and the *hfq*$^{\mathrm{FLAG}}$ strain (Hfq-coIP). The Hfq-associated RNA was converted to 5'monophosphate RNA by treatment with tobacco acid pyrophosphatase (TAP). Upon 5'-linker ligation and poly(A) tailing RNA was converted into cDNA, and a total of $\approx$175,000 cDNAs for each library pooled from two independent biological experiments was then characterised by high-throughput pyrosequencing (Margulies *et al.*, 2005).

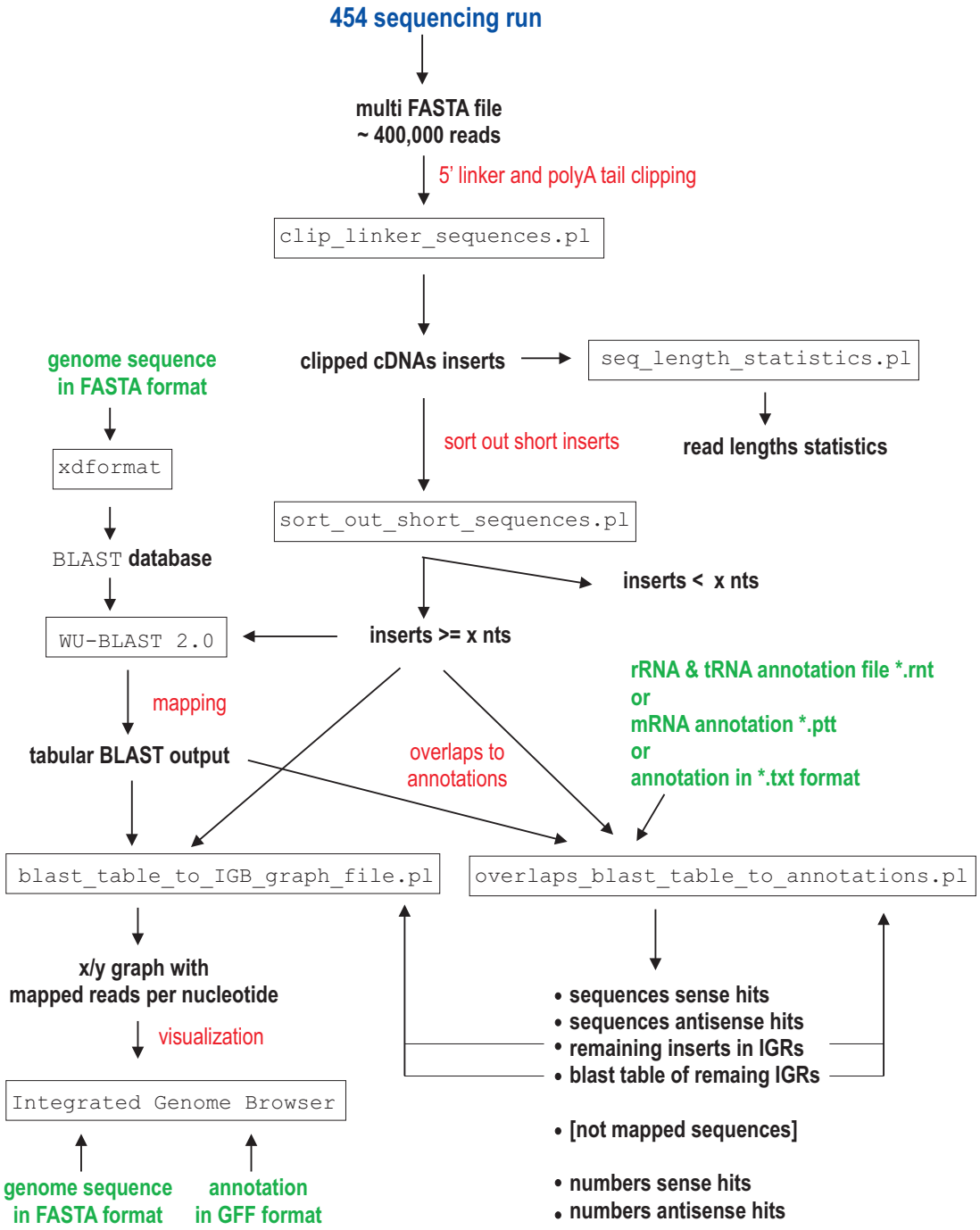


Figure 5.2: Work flow for the analysis of 454 sequencing data. Several Perl scripts have been developed for the analyses of 454 sequencing data. The work-flow including input files and generated output files is shown. The general analysis steps are indicated in red, genome sequence or annotation files in green, and output files in black. The developed Perl scripts are framed in black.

5.1.2. Biocomputational analysis of 454 data

After sequencing different libraries together in a 454 sequencing run, first the source of each read from the multiplex samples has to be determined by the help of specific barcode sequences that were attached to the 5' adaptor sequence during cDNA synthesis. Furthermore, 5'end linker sequences as well as poly(A)-tails have to be clipped from each read. After removal of short sequences, the next step in the bioinformatics analysis is alignment against a reference genome sequence, in this case *Salmonella enterica* subsp. enterica serovar Typhimurium str. LT2, as the SL1344 sequence is not annotated yet. Finally, visualization of mapped reads in an appropriate genome browser as well as determining how many reads overlap annotated genes are subsequent steps in the 454 sequencing analysis pipeline.

For this purpose, several Perl[1] (Practical Extraction and Report Language) scripts were developed. Figure 5.2 gives an outline of the analysis workflow. The scripts are explained in more detail in the following sections.

5.1.2.1. Removal of 5' end linker sequences and poly(A) tails

During 454 sequencing, several libraries, that can be distinguished by their specific 5' linker tags, were mixed together. From this mixed 454 run, reads for the Hfq-coIP library (Hfq-coIP) and a control library (coIP-Ctr) could be identified via their specific 5' linker 4-mer barcode tags: 'CGCA' and 'GCTC', respectively.

A typical 454 read looks as follows:

```
      tag     5'-adapter     cDNA insert (1-800nt)     polyA-tail       3'-adapter
5'-CTNNNNGACCTTGGCTGTCACTCANNNNNNNNN...NNNNNNNNNNNAAAAAAA...AAAAAAAGCGGGGCGATGTCTCGT-3'
```

The four 'N' nucleotides indicated in bold correspond to the 4-mer barcode tag which was attached to the 5' end linker and which is specific for each library. For long inserts, reads are often not long enough to reach the poly(A) tail and 3'-adaptor due to 454 sequencing length limitations. Thus, reads can either cover the full 3'-end adaptor or end at any point within the insert, poly(A)-tail or 3'-adapter.

To sort out reads for a specific library from a 454 run with mixed libraries, a Perl script `clip_linker.pl` was written which extracts reads for a given tag and in the same step clips the 5'end linker and poly(A) tail sequences. The usage is as follows:

[1] `www.perl.org`

```
Usage:

clip_linker_seq.pl extracts sequences with a specific 4-mer linker tag
out of a multi-FASTA file from a 454 run with mixed libraries.
In addition 5'end linker sequences and poly(A) tails will be clipped.
If the option -l all is specified, all sequences will be clipped and
not only those with a specific linker tag.

clip_linker.pl -f 454_sequence_file.fasta -l linker_tag (4mer) [-p prefix outfile]
              [-o] [-t]

Options:
    -f  454_sequence_file.fasta
    -l  linker_tag (4mer)
   [-p  prefix outfile]
   [-o] clips only 5'end linker sequences
   [-t] clips only poly(A) tail
   [-h] print this message
```

The script generates an outfile `*.inserts` in plain Vienna format with all clipped inserts of reads that contained the user defined tag. In case only one library was sequenced in the 454 run, the barcode identification is not necessary and all sequences can be clipped by the option `-l All`. Identification of 5'end linker and poly(A) tails is achieved by a pattern search (Dsouza *et al.*, 1997). For example, for the specified tag 'CGCA', `clip_linker.pl` calls the pattern matching program Patscan[2] with the following nucleotide pattern:

```
%Patscan Pattern for 454 sequencing data
CTCGCA
GACCTTGGCTGTCACTCA[1,1,1]
1...3000
( (AAAAA 1...15 AAA[1,0,0] (G | 0...0) CGGGGCGATGTCTCG[1,1,2] | AAAAA AAAAAAAAA[1,0,0] )
\ | ( AAAAA AAAAAAAAA[0,0,1] | AAAAA AAAAAAAAA[1,1,0]))
```

The user specified 4-mer tag is indicated in bold and will be set during each run of the script according to the user defined sequence. The first part of the pattern recognizes the 5' linker, the middle part the insert, and the last part defines several alternative patterns for recognition of the poly(A) tail alone or in combination with the 3' adapter. As mentioned above, some reads do not reach the poly(A)-tail. To recognize also these reads with the above specified pattern, an artifical poly(A)-tail A_{25} is added to every read before the pattern search is started. Thus, for reads without poly(A)-tail the whole read down to its 3'end will be taken after clipping.

5.1.2.2. Read length distribution

To get an overview of the cDNA insert lengths distribution for a single library or a complete run, the script `lengths_statistics.pl` counts the observed read lengths for a given insert file. It is used as follows:

[2] `http://www-unix.mcs.anl.gov/compbio/PatScan/HTML/patscan.html`

```
Usage:

seq_length_statistics.pl counts the read length for each sequence of a given insert file.
In addition it generates a text output file with the corresponding insert length
distribution.

statistics_for_sequences.pl -f inserts_file
Options:
    -f  inserts_file
   [-h] print this message
```

It produces two text output files. The first one (`*_lengths.txt`) lists for each read the actual sequence number, name, length, and sequence:

```
Number Name Length Sequence
1 >000021_2729_0305:[6,67]  27 TTCCAGAGTTCGAATCTCTGCGAACTT
2 >000011_2713_0289:[6,70]  30 GGTGAGGTGTCCGAGTGGCTAGAAGGAGAC
3 >000012_2717_0300:[6,108] 67 TTAGGCATTAACGGGAACCGGAGCGTTCCCGATTCACCATGGATGGCCTTTTCGG
4 >000015_2740_3628:[6,105] 66 TAAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTAGGCGGC
...
```

The second one (`*_distribution.txt`) contains the read length distribution:

```
Total number of sequences: 15384

Sequence_length Number_of_sequences Per_mille_reads_of_library
1   5     0.3250130005
2   1     0.0650026001
3   2     0.1300052002
4   16    1.0400416017
5   74    4.8101924077
6   94    6.1102444098
7   144   9.3603744150
8   210   13.6505460218
...
```

5.1.2.3. Removal of short reads

As very short sequences can not be specifically mapped to the genome, a script which sorts out the sequences according to a user defined threshold *l* was developed. The usage is as follows:

```
Usage:

Sort_out_short_sequences.pl sorts out sequences with a length shorter than
a user defined threshold
Sort_out_short_sequence.pl -f inserts_file -l length cut-off[nt]
Options:
    -f  inserts_file
    -l  length cut-off[nt]
   [-h] print this message
```

For the analysis of the *Salmonella* Hfq-coIP and coIP-Ctr libraries, a threshold of 18 bp was selected based on some initial *BLAST* test searches with reads of different lengths against the *Salmonella* genome. The script `Sort_out_short_sequence.pl` produces two output files of which `*_less_18nt.inserts` contains all sequences with a length shorter than the user defined cut-off l (here 18 bp) and `*_min_18nt.inserts` all sequences with a length $\geq l$, respectively.

5.1.2.4. Read mapping to a reference genome

After linker clipping and removal of short reads, the sequence reads have to be aligned to a reference sequence, usually a reference genome. Here, the best-known pairwise alignment algorithm, `BLAST` (Basic Local Alignment and Search Tool) was selected for mapping. It uses a seed-and-extend strategy, whereby alignments that are extended form short, identical seed substrings. Although this approach does not work well for short query sequences that contain mismatches, it works quite well for alignment of 454 sequencing results as these produce on average longer reads compared to other sequencing methods like Solexa[3] and Solid[4].

Here, all reads $\geq$18 nt were mapped to the *Salmonella* genome (`NC_003197`) using `WU-Blast`[5] 2.0 with the following parameters:

```
% blastn database *_min_18nt.inserts -B=1 -V=1 -m=1 -n=-3 -Q=3 -R=3
\ -gspmax=1 -hspmax=1 -mformat=2 -e=0.0001
```

which have the following meanings:

- $-B = 1$ maximal number of database sequences to report alignments from
- $-V = 1$ maximal number of reported alignments for a given database sequence
- $-m = 1, -n = -3$ match, mismatch scoring system
- $-Q = 3$ penalty score for a gap of length 1
- $-R = 3$ penalty score for extending a gap by each letter after the first
- $-gspmax = 1$ max. number of gapped high-scoring pairs (GSPs) saved per subject sequence (default 0; 0 $\rightarrow$ unlimited)
- $-hspmax = 1$ max. number of ungapped HSPs saved per subject sequence (default 1000; 0 $\rightarrow$ unlimited)
- $-mformat = 2$ specifies tabular output format (default 1)
- $-e = 0.0001$ Expectation value (E) [Real] default = 10.0

[3] `www.illumina.com`

[4] `solid.appliedbiosystems.com`

[5] `http://blast.wustl.edu/`

Note that with these settings only the highest-scoring hit (first hit in the BLAST table) was reported. This means that for reads matching to multiple locations in the genome only the first hit was considered. As a read corresponds to a transcript which can only derive from one location of the genome, this approach was taken to simplify the counting of mapped reads and overlaps to annotations later on.

5.1.2.5. Overlaps to annotations

Next, the questions how many reads overlap to annotated genes and what are the fractions of *e. g.* rRNA, tRNA, mRNA etc. in the Hfq-coIP and coIP-Ctr libaries were addressed. Therefore, a script called `overlaps_blast_table_to_annotations.pl` with the following usage was generated:

```
Usage:

overlaps_blast_table_to_annotations.pl sorts out any specified annotated genes like
tRNAs and rRNAs from tabular WU-Blast output file and also sequences with no blast
hit if the option -n is specified.
The blast input file has to be a WU-Blast output file in tabular format (-mformat=2).
The annotation file can be either in *.rnt or *.ptt format and has to be specified by
the option -a annotation.rnt/ptt OR in tabular text format and has to be specified by
-t annotation.txt.
The tabular format contains the following fields separated by tab:

Product_name Start End Strand Length GeneID Locus Locus_tag Links

Sort_out_annotated_seqs_from_454_sequences.pl
Options:
      -i   insert_file
      -b   WU-Blast_file in tabular format (generated by mformat = -2)
      -a   annotation file in *.rnt (rRNA and tRNA) or *.ptt (mRNA) format with gene
           positions of genes that should be sorted out
           !OR! option:
      -t   annotation file in *.txt format with gene positions of genes that should be
           sorted out
     [-n]  write outfile with sequences that have no blast hit
     [-l]  write also all gene names with 0 overlaps in sense hits
     [-p   overlap_length (nt)] allows partial overlap for sorting out, p defines
           minimum required overlap
     [-h]  print this message
```

The script requires an input file with inserts, *e. g.*, `*_min_18nt.inserts`, which is used as input for the WU-Blast mapping, and a second input file with the `BLAST` results in tabular format (generated by the option $-mformat = 2$ by `WU-Blast`) `*.blastn`. In addition, a file with annotations of genes that should be used for the overlap has to be provided. Annotations of tRNAs, rRNAs, hkRNAs (housekeeping RNAs), and several ncRNAs for *Salmonella* can be downloaded from NCBI[6] and are listed in the file `NC_003197.rnt`. Annotations for proteins can be found in the file `NC_003197.ptt`, respectively. Annotations in the `*.rnt` and `*.ptt` format can be passed by the option $-a\ annotationfile$. In addition, own annotations can be provided as a text

[6] `ftp://ftp.ncbi.nih.gov/genomes/Bacteria`

file and passed to the program by the option $-t\ annotationfile$. For the `*.txt` annotation file, the following tab delimited text format is required:

Product Name	Start	End	Strand	Length	GeneID	Locus	Locus_tag	Links
16S ribosomal RNA	289189	290732	+	1544	1251767	rrsH	STM0249	-
Ile tRNA	290800	290873	+	74	1251768	ileV	STM0250	-
Asx tRNA	290986	291058	+	73	1251769	alaV	STM0251	-
Met tRNA	738643	738716	-	74	1252199	metT	STM0679	-
Lys tRNA	818775	818847	+	73	1252271	lysT	STM0751	-
...								

For the *Salmonella* Hfq libraries, three files with annotations were compiled in this format: (1) rRNAs, tRNAs, and hkRNAs (`LT2_rRNA_tRNA_hkRNA.txt`), (2) known sRNAs (`LT2_known_sRNAs.txt`), and (3) predicted sRNAs (`LT2_predicted_sRNAs.txt`), and are provided in the Appendix (Tables 10.5, 10.6, and 10.7). Furthermore, a user defined cut-off p has to be specified. The script counts a read x as overlapping to an annotation y if the read is located within the annotation or at least p nucleotides of the read x overlap with the annotation y. The script generates four main output files:

1. `*_pPnt_annotation_file.sense_hits`
   contains all inserts that are located sense to the annotations in the given annotation file and overlap the annotation for at least P nucleotides specified by the option $-p$.

2. `*_pPnt_annotation_file.antisense_hits`
   contains all inserts that are located antisense to the annotations in the given annotation file and overlap the annotation for at least P nucleotides specified by the option $-P$.

3. `*_pPnt_annotation_file.igr`
   contains all inserts that do not overlap the annotations that have been specified.

4. `*_pPnt_annotation_file_igr.blast`
   contains the blast results in tabular format for all inserts that do not overlap the annotations that have been specified.

In the first two files, overlaps to annotated regions are sorted in decreasing order of number of overlapping reads. That means, annotations with the highest numbers of overlapping reads are listed first and also the total number of reads that overlap the annotation is given in the first line of each overlap. In case the option $-n$ is specified, an additional output file (`*.not_hit`) will be generated which contains all inserts for which no hit in the BLAST table is provided. The two files `*.igr` and `*.blast` with the remaining BLAST hits and inserts can then be used as input files for a new round of overlaps with the next annotation file and so forth. This is helpful when, for example, several annotations overlap each other at certain genome positions. Furthermore, certain classes of RNAs can be sorted stepwise, *e. g.*, first very abundant RNAs like rRNAs, tRNA, next mRNAs, and finally sRNAs. For the *Salmonella* libraries, first overlaps to rRNA, tRNA and hkRNAs were made, then for known sRNAs, followed by mRNA annotations, and finally annotations for predicted sRNAs in *Salmonella*. To facilitate the extraction of overlap numbers, two more output files are generated,

`*sense_hits.numbers` and `*antisense_hits.numbers`, which summarize the overlaps in the following format:

```
Sense overlap numbers for Hfq_CoIP_mixed_min_18nt_p20_LT2_rRNA_tRNA_hkRNA.txt
STM0249 11209 rrsH
STM0252 11091 rrlH
STM1944 6200  glyW
STM2411 3488  alaX
STM3934 2296  proM
STM0253 2285  rrfH
...
```

The first column indicates the gene name, the second the number of overlapping reads, and the third column the alternative gene name or description, respectively. The option $-l$ writes also annotations with zero overlaps to the output file which facilitates comparison of overlap numbers if, *e. g.*, the numbers of different libraries should be copied into one table.

5.1.2.6. Visualization of mapped sequences

For visualization of the location of blast hits, graph files were calculated that can be loaded into the Integrated Genome Browser (IGB)[7] of Affymetrix. This browser allows the visualization of both whole genomes and individual genomic regions. Graphs representing the number of mapped reads were calculated from `BLAST` tables and loaded together with the sequence (`NC_003062.fa`) and annotation (`NC_003062.gff`) of the *Salmonella* chromosome. The script `blast_table_to_IGB_graph_file.pl` can be used as follows:

```
Usage:

blast_table_to_IGB_graph_file.pl generates an x/y graph file for visualization in the
Integrated Genome Browser (Affymetrix) from a tabular WU-Blast output file (-mformat=2).
For each read in the blast table that was mapped to position x to y in the genome
it will add +1 from position x to y if the read was mapped on the plus strand, and -1
if it was mapped on the minus strand, respectively.
If the option [-n NF] is used, the mapped reads per nucleotide position will be divided
by the value of the normalization factor NF and multiplied by 1000. Thus, the y values
are given as ''per mill'' reads of the given normalization factor. Per default, NF is set
to one and the y-values correspond to the actual number of mapped reads per nucleotide.

blast_table_to_IGB_graph_file.pl -b blast_table -o outputname
                                 [-n normalization_factor (default = 1)]
Options:
   -b  blast_table (generated by Wu-Blast with option -mformat=2)
   -o  outputname
  [-n  normalization_factor (default = 1)]
  [-h] print this message
```

It generates the two output files, `*_NF_x_plus.gr` and `*_NF_x_minus.gr`, which contain a list of x / y coordinates for the plus (`*_plus.gr`) and the minus (`*_minus.gr`) strand. The x coordinate corresponds to the nucleotide position in the genome, the y coordinate to the number

[7] `http://www.affymetrix.com/support/developer/tools/download_igb.affx`

of mapped reads at this position x. The files were generated by iterating through all reads in the BLAST table and adding +1 (-1 for hits on the minus strand) at each nucleotide position x to y in the chromosome to which the read was mapped from position x to y by WU-Blast. Thus, the *_plus.gr visualizes BLAST hits on the plus strand, the *_minus.gr BLAST hits on the minus strand, respectively. To compare different libraries, a normalization factor (NF), *e. g.* the total number of mapped reads for a given library, can be passed to the script by the option $-n$. If so, the number of mapped reads per nucleotide position is divided by the normalization factor n and multiplied by 1000. This is necessary in case libraries have big differences in the number of sequenced reads. The normalization factor used for the calculation of the y-value is indicated in the output file name '...NF_x_...'.

The graphs described above contain x/y-values only at positions in the genome where at least one read was mapped. In the IGB, it is possible to apply several mathematical operations to graph files. However, to divide the values of one graph by the values of a second graph to calculate enrichment factors, the graphs must have the same number of, and equal, x-values. Therefore, a slightly modified script blast_table_to_complete_IGB_graph_file.pl was generated which sets the regions where no read was mapped to "+1" for the plus strand and to "-1" for the minus strand, respectively. The usage is as follows:

```
Usage:

blast_table_to_complete_IGB_graph_file.pl generates an x/y graph file for visualization in
the Integrated Genome Browser (Affymetrix) from a tabular Wu-Blast output file
(-mformat=2).
For each read in the blast table that was mapped to position x to y in the genome
it will add +1 from position x to y if the read was mapped on the plus strand, and -1 if
it was mapped on the minus strand, respectively.
Genome positions where no read was mapped will be set to 1 (-1 for the minus graph) to
allow mathematical operations between two graphs in the IGB which requires the same number
of x coordinates for the two graphs.
If the option [-n NF] is used, the mapped reads per nucleotide position will be divided
by the value of the normalization factor NF and multiplied by 1000. Thus, the y values are
given as ''per mill'' reads of the given normalization factor. Per default, NF is set to
one and the y values correspond to the actual number of mapped reads per nucleotide.

blast_table_to_complete_IGB_graph_file.pl -b blast_table -o outputname -l genome_length
                                          [-n normalization_factor (default = 1)]
Options:
   -b  blast_table (generated by Wu-Blast with option -mformat=2)
   -o  outputname
   -l  genome_length
  [-n  normalization_factor (default = 1)]
  [-h] print this message
```

5.1.3. Analysis of cDNA sequencing results of Hfq-associated RNA

For both libraries, Hfq-coIP and coIP-Ctr, $\approx$ 175,000 cDNA reads were sequenced (see Table 5.1). The resulting sequences for the Hfq-coIP, from here on referred to as 'Hfq cDNAs', ranged in length from 1 to 145 bp, and 92% were $\geq$ 18 bp (Fig. 5.3 A). Disregarding small cDNAs ($<$18 bp), 122,326

Table 5.1: Read distribution of Hfq-coIP and coIP-Ctr libraries. The total number of sequenced reads for the Hfq-coIP and coIP-Ctr libraries as well as fractions of reads $<$ or $\geq$ 18 bp are given. Furthermore, the number of mapped reads (BLAST hits) for sequences $\geq$ 18 bp and the number of reads overlapping to annotated genes, known and predicted sRNAs as well as intergenic regions (IGR) are listed.

	Hfq-coIP	**coIP-Ctr**
total reads	176,907	175,142
$<$ 18 bp	13,736	12,069
$\geq$ 18 bp	163,171	163,073
BLAST hits	122,326	145,873
$\geq$ 18 bp, but no hit	40,845	17,200
rRNA, tRNA, hkRNA[a]	57,529	132,148
known sRNAs	11,922	1445
mRNA	34,136	7911
predicted sRNAs	647	139
antisense mRNA	2174	342
antisense ncRNA	42	7
IGR	15876	3881

[a] This set includes the housekeeping RNAs (hkRNAs) tmRNA, M1 RNA, and SRP RNA.

sequences could be unequivocally mapped to the *Salmonella* genome by WU-BLAST[8](Table 5.1). About half of the mapped cDNAs (57,529) were derived from rRNA, tRNA, and housekeeping RNAs (tmRNA, M1 RNA, and SRP RNA; Fig. 5.3 B). Of the remaining 64,797 sequences, the majority corresponded to mRNA regions (53% matched the sense strand of protein-coding regions), followed by known/predicted conserved sRNAs (18%; Hershberg *et al.*, 2003; for distribution see Fig. 5.3 C), predicted *Salmonella*-specific sRNAs (1%; Pfeiffer *et al.*, 2007) and sequences that were antisense to ORF regions (3%). The remaining 25% of cDNAs mostly represented intergenic regions (IGRs) and 5' or 3' UTRs, with a few antisense transcripts to tRNAs, rRNAs, and sRNAs (0.1%; Fig. 5.3 B).

To confirm that the procedure did effectively enrich Hfq-associated RNAs, the 175,142 cDNAs from the control coIP using wild-type *Salmonella* (expressing untagged Hfq) were analysed. Of these 'Control cDNAs' which ranged in length from 1 to 290 bp (Fig. 5.3 A), 145,873 sequences were $\geq$18 bp in size and could be mapped to the *Salmonella* chromosome. Most of the inserts (91%) were abundant rRNA, tRNA, and housekeeping RNA transcripts (Fig. 5.3 B). The remaining 13,725 sequences were used to calculate the level of enrichment of Hfq-bound RNA (see below).

5.1.4. Visualizing Hfq-dependent RNAs at the nucleotide level

Upon WU-BLAST matching, the number of cDNA hits for each nucleotide position for either strand of the *Salmonella* chromosome was calculated using the Perl script `blast_table_to_IGB_graph_file.pl` and visualized using the Integrated Genome Browser (IGB, Affymetrix). Figure 5.4 shows the distribution of cDNA sequences over a subsection of the

[8] `http://blast.wustl.edu/`

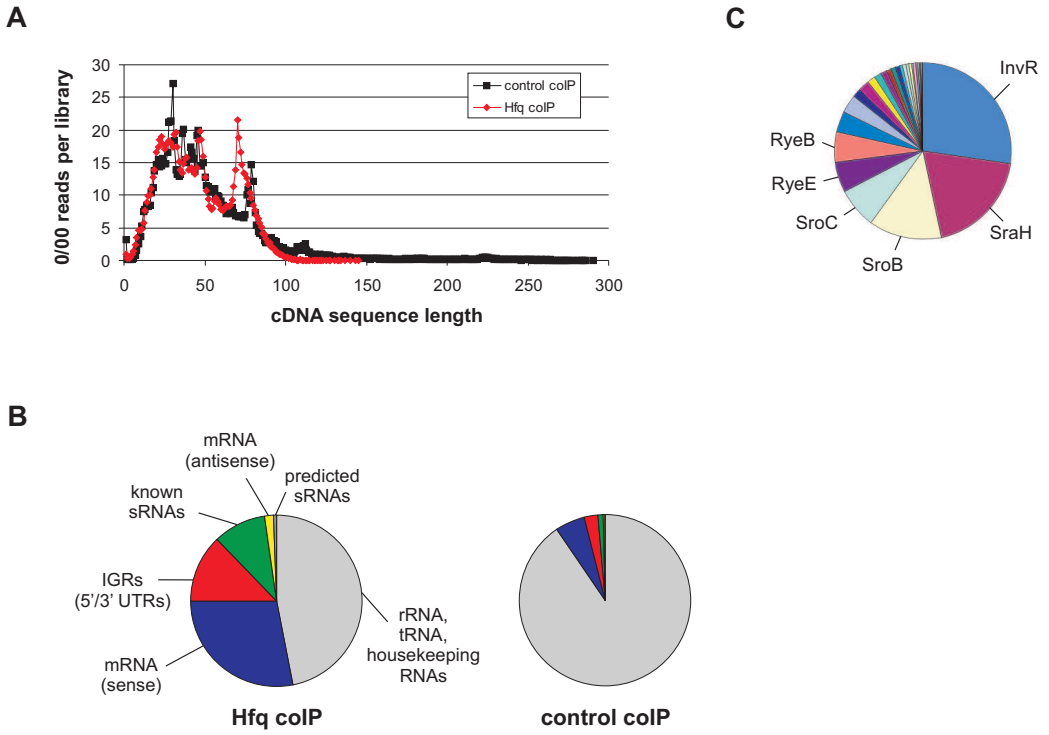


Figure 5.3: Statistical analysis of the cDNA sequencing results of Hfq-associated RNA. (A) The pyrosequencing results were analysed by plotting the number of cDNA reads *vs* the length of the clipped reads in bp. The length distribution of all resulting sequences is shown. **(B)** Pie diagram of the different RNA species contained in all sequences that mapped to the *Salmonella* genome. The rRNA, tRNA and housekeeping RNAs are shown in grey. Left panel: Hfq-coIP, right panel: control coIP. **(C)** Pie diagram for all Hfq-associated sequences that unequivocally mapped to known sRNA sequences. The names of the six most strongly recovered sRNAs are given. (Figure was adapted from Sittka *et al.*, 2008.)

genome, *i. e.* the ≈40 kb SPI-1 virulence region, for which a strong enrichment of Hfq cDNAs over the Control cDNAs was observed. As well as the 35 mRNAs of protein-coding genes, SPI-1 encodes the Hfq-dependent InvR sRNA (Pfeiffer *et al.*, 2007). SPI-1 represents an example of an entire genomic region highly enriched in the Hfq-coIP library. In contrast, very few cDNA sequences mapping to SPI-1 are contained in the coIP-Ctr library. The flanking genes of *invR* (i.e. the border of SPI-1) give a nice example of the specificity of the method (Fig. 5.4). While cDNAs mapping to the InvR sRNA gene represent the most abundant cluster in the Hfq-coIP library, the genes in the closest proximity are barely represented in this library. In addition, the example of InvR underlines the reliability of the method to identify Hfq-dependent sRNAs. Enrichment of InvR by coIP with FLAG-tagged Hfq was previously demonstrated by Northern blot analysis (Pfeiffer *et al.*, 2007), and this result is confirmed by the strong cDNA peak seen at the *invR* locus located at the right-hand SPI-1 border (Fig. 5.4).

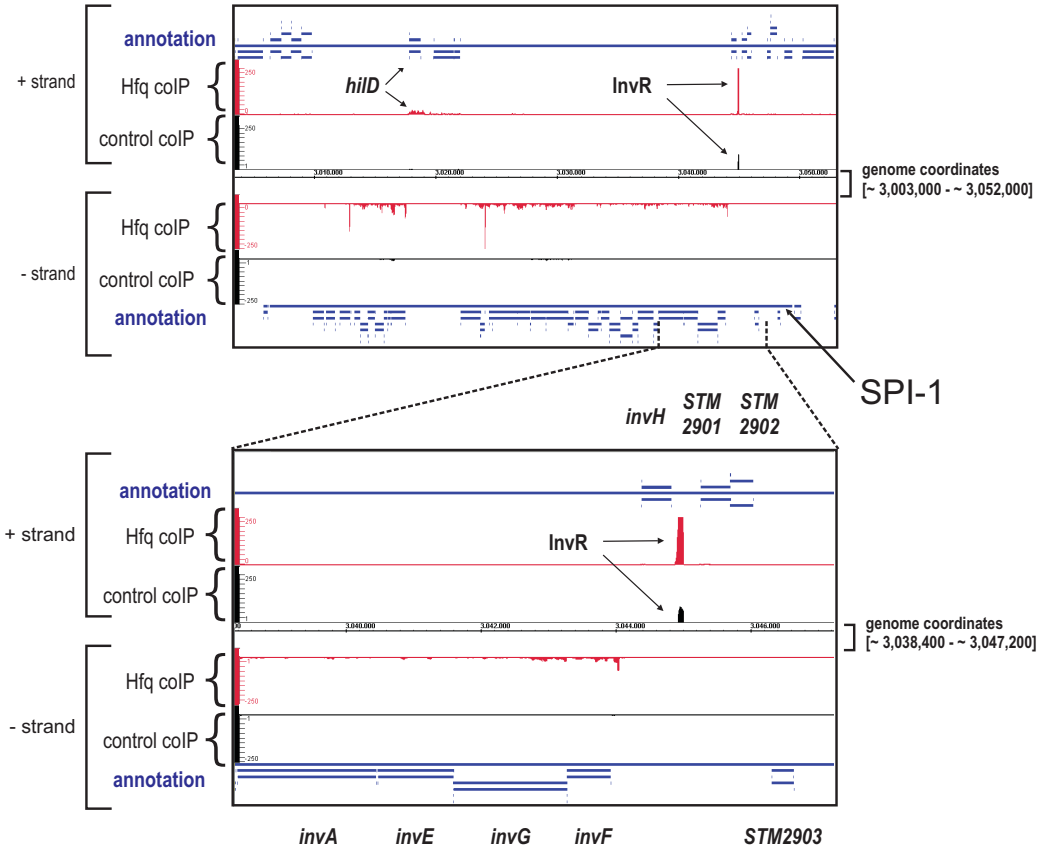


Figure 5.4: Visualization of pyrosequencing data for the *Salmonella* pathogenicity island 1 (SPI-1). The upper panel shows an extraction of the screenshot of the Integrated Genome Browser, with the mapped Control and Hfq cDNAs of the SPI-1 region. Shown are the annotations for the (+) and (-) strand (blue), the cDNA sequence distribution from the Hfq-coIP for the (+) and (-) strand (red), the cDNA-clone distribution for the control coIP for the (+) and (-) strand (black), and the genome coordinates in the center for the entire SPI-1. The annotation for SPI-1 and the Hfq-coIP peaks for *hilD* and the sRNA InvR in the Hfq-coIP are indicated. Note that the clone numbers per nucleotide are scaled to a maximum of 250 for the Hfq and the control coIP, which truncates the high peak for InvR in the Hfq-coIP library (>3,000 cDNAs). The lower panel shows a close-up of the *invR* locus and its adjacent genes. (Figure was adapted from Sittka *et al.*, 2008.)

5.1.5. Hfq-dependent sRNAs are highly associated with Hfq

Inspection of the cDNA libraries revealed that a major class of reads was derived from sRNA regions. These sRNAs, as well as their enrichment by Hfq coIP, are listed in Table 5.2 and Table 10.8 in the Appendix. The three most abundant sRNAs according to the numbers of Hfq cDNA sequences are InvR, SraH (a.k.a. RyhA) and SroB (RybC) and are known to be strongly bound by Hfq (Pfeiffer *et al.*, 2007; Zhang *et al.*, 2003); coIP of Hfq enriched these three sRNAs by 30- to 57-fold in comparison to the control set. For example, InvR, which binds Hfq with a K_D of 10 nM (Pfeiffer *et al.*, 2007), was represented by 3,236 Hfq cDNAs and 113 Control cDNAs (see Table 5.2). In contrast, other sRNAs not expected to be Hfq-dependent were found in equal numbers in the two samples, *e. g.*, the CsrB or CsrC sRNAs which target the conserved RNA-binding

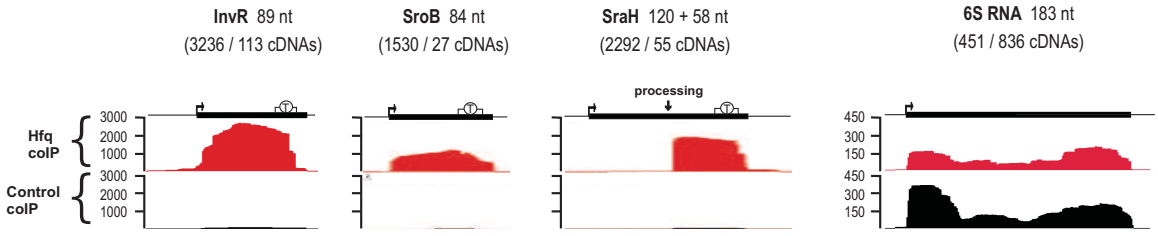


Figure 5.5: Visualization of the clone distribution of exemplary Hfq dependent and independent sRNAs in *Salmonella*. Clone distribution for sequences mapped to InvR, SroB, SraH, or 6S sRNAs (red: Hfq-coIP, black: control coIP). The vertical axis indicates the number of cDNA sequences that were obtained. A bent arrow indicates the sRNA promoter, a circled 'T' its transcriptional terminator. (Figure was adapted from Sittka *et al.*, 2008.)

protein CsrA (Babitzke & Romeo, 2007) were represented by almost equal numbers in the Hfq and Control cDNAs (CsrB, 67/69; CsrC, 63/64; Table 5.2). Moreover, cDNAs of the abundant yet Hfq-independent 6S RNA (Wassarman *et al.*, 2001) were found in smaller numbers in the Hfq than in the control library (451 versus 836; Table 5.2).

Fig. 5.5 illustrates the distribution of cDNAs of the three predominant Hfq-bound RNAs and of the Hfq-independent 6S RNA. The cDNAs of both the InvR (Pfeiffer *et al.*, 2007) and SroB (Vogel *et al.*, 2003) sRNAs mapped along the entire RNA coding sequence from the transcriptional start site to the Rho-independent terminator. SraH, which is transcribed as an unstable 120 nt precursor and is processed into an abundant $\approx$58 nt RNA species (3' part of SraH; Argaman *et al.*, 2001; Zhang *et al.*, 2003), was almost exclusively recovered as the processed sRNA. Notably, the borders of the cDNA clusters were in perfect agreement with previous 5' and/or 3' end mapping data of the four sRNAs (Argaman *et al.*, 2001; Brownlee, 1971; Pfeiffer *et al.*, 2007; Vogel *et al.*, 2003). In other words, the present cDNA sequencing approach not only detects association with Hfq but also identifies the termini of expressed sRNAs at nucleotide-level resolution.

5.1.6. Identification of expressed *Salmonella* sRNAs

To evaluate the sRNA expression profile of *Salmonella* more extensively, three classes of sRNA candidate loci were analysed for coverage by the Hfq and Control cDNAs. First, cDNAs of *E. coli* sRNA candidate loci with predicted conservation in *Salmonella* were inspected (Argaman *et al.*, 2001; Hershberg *et al.*, 2003; Kawano *et al.*, 2005a; Wassarman *et al.*, 2001; Rivas *et al.*, 2001; Vogel *et al.*, 2003; Zhang *et al.*, 2003). Second, cDNAs of *Salmonella*-specific sRNAs predicted in two recent global screens (Padalon-Brauch *et al.*, 2008; Pfeiffer *et al.*, 2007) were counted. Third, cDNAs from one third of the *Salmonella* chromosome (first 1.6 Mb) and all major five pathogenicity islands were manually inspected for expression patterns of IGRs indicative of new sRNA genes, and for possible enrichment by Hfq coIP. Of the latter two classes of candidates (summarized in Table 10.8 in the Appendix), those with an Hfq enrichment factor $\geq$10 and/or candidates showing good promoter/terminator regions, were selected for Northern blot analysis.

Table 5.2: Compilation of expressed *Salmonella* sRNAs and their enrichment by Hfq-coIP.

sRNA[1]	Alternative IDs[2]	Reference[3]	Adjacent genes[4]	Orientation[5]	Start[6]	End[6]	Reads coIP-Ctr[7]	Reads Hfq-coIP[8]	Enrichment[9]	Northern blot[10]
sgrS	*ryaA*	I	*yabN/leuD*	←→←	128574	128812	3	61	20.3	
isrA	-	II	STM0294.ln/STM0295	→→→	339338	339760	0	0		
sroB	*rybC*	I	*ybaK/ybaP*	←→←	556005	556085	27	1530	56.7	
sroC	-	I	*gltJ/gltI*	←←←	728913	728761	26	898	34.5	
rybB	p25	III	STM0869/STM0870	→←←	942632	942554	3	103	34.3	
STnc490[11]	-	IV	*clpA/tnpA*_1	→←→	1024975	1025165	75	385	5.1	≈85nt
isrB-1	-	II	*sbcA*/STM1010	←→←	1104179	1104266	2	4	2.0	
STnc500	-	IV	STM1127/STM1128	←←←	1216157	1216440	7	84	12.0	≈65nt
STnc150	-	V	*icdA*/STM1239	→←→	1325914	1325649	0	1	≥1.0	≈90nt
isrC	-	II	*envF/msgA*	←→←	1329145	1329432	0	1	≥1.0	
STnc520	-	IV	STM1248/STM1249	→←←	1332809	1334044	12	100	8.3	≈80nt
isrD	-	II	STM1261/STM1263	→←→	1345788	1345738	0	0		
ryhB-2	*isrE*	II	STM1273/*yeaQ*	→←→	1352987	1352875	0	0		
STnc540	-	IV	*himA/btuC*	→→→	1419369	1419570	7	23	3.3	≈85nt
rprA	IS083	I	*ydiK/ydiL*	←←←	1444938	1444832	37	286	7.7	
rydB	tpe7, IS082	I	*ydiH*/STM1368	→→←	1450415	1450519	4	10	2.5	
STnc570[12]	*yneM* ORF	IV	*ydeI/ydeE*	→←←	1593723	1594413	2	21	10.5	≈190nt
STnc560		IV	*ydeI/ydeE*	→→←	1593723	1594413	10	290	29.0	≈90nt
isrF	-	II	STM1552/STM1554	→←←	1630160	1629871	1	0		
rydC	IS067	I	STM1638/*cybB*	→→←	1729673	1729738	5	245	49.0	
micC	IS063, tke8	III	*nifJ/ynaF*	→←→	1745786	1745678	0	15	≥15.0	
STnc580	-	IV	*dbpA*/STM1656	←←←	1749662	1750147	11	311	28.3	≈100nt
ryeB	tpke79	I	STM1871/STM1872	→←←	1968155	1968053	24	653	27.2	
dsrA	-	I	*yodD/yedP*	→←→	2068736	2068649	6	149	24.8	
rseX	-	I	STM1994/*ompS*	←→→	2077175	2077269	0	3	≥3.0	
ryeC	tp11	I	*yegD*/STM2126	→→→	2213871	2214016	42	72	1.7	
cyaR	*ryeE*	III	*yegQ*/STM2137	→→→	2231130	2231216	31	659	21.3	
isrG	-	II	STM2243/STM2244	←→→	2344732	2345013	0	0		
micF	-	III	*ompC/yojN*	←→→	2366913	2367005	0	11	≥11.0	

continued on next page

sRNA[1]	Alternative IDs[2]	Reference[3]	Adjacent genes[4]	Orientation[5]	Start[6]	End[6]	Reads coIP-Ctr[7]	Reads Hfq-coIP[8]	Enrichment[9]	Northern blot[10]
isrH-2	-	II	*glpC*/STM2287	→←→	2394582	2394303	0	0		
isrH-1	-	II	*glpC*/STM2287	→←→	2394753	2394303	0	0		
STnc250[12]	*ypfM* ORF	V	*acrD/yffB*	→←→	2596882	2596789	6	24	4.0	≈ 220nt
ryfA	tp1	I	STM2534/*sseB*	→→←	2674934	2675228	3	6	2.0	
glmY	tke1, *sroF*	I	*yfhK/purG*	←←←	2707847	2707664	20	92	4.6	
isrI	-	II	STM2614/STM2616	→←←	2761576	2761329	0	2	≥2.0	
isrJ	-	II	STM2614/STM2616	→←←	2762031	2761957	1	0		
isrK	-	II	STM2616/STM2617	←←←	2762867	2762791	0	0		
isrB-2	-	II	STM2631/*sbcA*	→←→	2770965	2770872	0	0		
isrL	-	II	*smpB*/STM2690	→←→	2839399	2839055	0	0		
isrM	-	II	STM2762/STM2763	←→→	2905050	2905378	0	0		
isrN	-	II	STM2764/STM2765	←→←	2906925	2907067	0	0		
micA	*sraD*	I	*luxS/gshA*	←→←	2966853	2966926	1	128	128.0	
invR	STnc270	III	*invH*/STM2901	→→→	3044924	3045014	113	3236	28.6	
csrB	-	III	*yqcC/syd*	←←←	3117059	3116697	69	67		
gcvB	IS145	III	*gcvA/ygdI*	←→←	3135317	3135522	12	402	33.5	
omrA	*rygB*	III	*aas/galR*	←←→	3170208	3170122	0	51	≥51.0	
omrB	t59, *rygA*, *sraE*	III	*aas/galR*	←←→	3170408	3170322	1	52	52.0	
STnc290	-	V	*tnpA_4*/STM3033	←←←	3194996	3194914	2	72	36.0	≈ 85nt
isrO	-	II	STM3038/STM3039	←→→	3198380	3198580	0	0		
ssrS	-	I	*ygfE/ygfA*	→→→	3222098	3222280	836	451		
rygC	t27	I	*ygfA/serA*	→→←	3222913	3223065	14	17	1.2	
rygD	tp8, C0730	I	*yqiK/rfaE*	→←←	3362474	3362327	17	104	6.1	
sraF	tpk1, IS160	I	*ygjR/ygjT*	→→→	3392069	3392261	0	25	≥25.0	
sraH	*ryhA*	I	*yhbL/arcB*	←→←	3490383	3490500	55	2292	41.7	
ryhB-1	*sraI*, IS176	I	*yhhX/yhhY*	←←→	3715495	3715401	0	2	≥2.0	
istR-1		VI	*ilvB/emrD*	←←→	3998147	3998018	0	0		≈ 75nt
istR-2		VI	*ilvB/emrD*	←←→	3998147	3998018	0	0		≈ 140nt
STnc400	-	V	STM3844/STM3845	→→→	4051145	4051340	112	42		≈ 55nt
glmZ	k19, *ryiA*, *sraJ*	I	*yifK/hemY*	→→←	4141650	4141854	20	196	9.8	
spf	*spf*	I	*polA/yihA*	→→←	4209066	4209175	2	33	16.5	

continued on next page

sRNA[1]	Alternative IDs[2]	Reference[3]	Adjacent genes[4]	Orientation[5]	Start[6]	End[6]	Reads coIP-Ctr[7]	Reads Hfq-coIP[8]	Enrichment[9]	Northern blot[10]
csrC	*sraK, ryiB,* tpk2	III	*yihA/yihI*	←→→	4210157	4210400	63	64		
isrP	-	II	STM4097/STM4098	←→←	4306719	4306866	0	2	≥2.0	
oxyS	-	I	*argH/oxyR*	→←→	4342986	4342866	0	10	≥10.0	
STnc620	-	IV	*ssb*/STM4257	→→→	4476817	4477856	4	41	10.3	nd
sraL	*ryjA*	III	*soxR*/STM4267	→←→	4505010	4504870	0	0		
STnc440	-	V	STM4310/*tnpA_6*	→→→	4559193	4559277	9	456	50.7	≈ 85nt
STnc460	-	V	STM4503/STM4504	→←→	4758332	4758187	0	0		np
isrQ	-	II	STM4508/STM4509	←→→	4762997	4763158	0	0		

[1] Gene names of *Salmonella* sRNAs identified in this and previous studies. The identification method is given in the third column. sRNA names follow the *Salmonella* and/or *E. coli* nomenclature referenced in Hershberg *et al.* (2003), Padalon-Brauch *et al.* (2008), and Papenfort *et al.* (2008).

[2] Alternative sRNA IDs. References in Hershberg *et al.* (2003), Padalon-Brauch *et al.* (2008), and Papenfort *et al.* (2008), except STnc490, 500, 520, 540, 560, 570, 580, which have been newly predicted in this study.

[3] Evidence for sRNAs in *Salmonella*. **(I)** Conserved sRNA found in *Salmonella* cDNA libraries, and expression previously shown in *E. coli* (relevant ref. in Papenfort *et al.*, 2008; Table 1). **(II)** sRNA previously predicted and validated on Northern blots in *Salmonella* by Padalon-Brauch *et al.* (2008). **(III)** sRNA previously validated on Northern blots in *Salmonella* (Altier *et al.*, 2000a; Figueroa-Bossi *et al.*, 2006; Fortune *et al.*, 2006; Papenfort *et al.*, 2006, 2008; Pfeiffer *et al.*, 2007; Sharma *et al.*, 2007; Viegas *et al.*, 2007). **(IV)** sRNA predicted through cDNA sequencing and validated on Northern blots in this study. **(V)** sRNA previously predicted by Pfeiffer *et al.* (2007) is recovered in cDNA sequences and validated on Northern blots in this study. **(VI)** IstR sRNAs (Vogel *et al.*, 2004) were not recovered in cDNA sequences but their expression in *Salmonella* was validated by Northern blot analysis in the complete study (Sittka *et al.*, 2008).

[4] Flanking genes of the intergenic region in which the sRNA candidate is located.

[5] Orientation of sRNA candidate (middle) and flanking genes on the clockwise (→) or the counterclockwise (←) strand of the *Salmonella* chromosome.

[6] Genomic location of sRNA candidate gene according to the *Salmonella typhimurium* LT2 genome. For STnc470 through STnc640 start and end of the entire intergenic region are given.

[7] Out of 145,873 sequences in total.

[8] Out of 122,326 sequences in total.

[9] Enrichment factor calculated by dividing the number of reads from Hfq-coIP by the number of reads from the coIP-Ctr.

[10] Denotes verification on Northern blot for new RNA transcripts; the estimated size is given in nucleotides (np = not probed; nd = no detectable transcript).

[11] The cDNA reads map antisense internally of the IS200 element. Based on sequence identity they map to all IS200 elements (*tnpA_1* to *tnpA_6*).

[12] STnc250 and STnc570 contain small ORFs annotated as *ypfM* and *yneM*, respectively, in *E. coli* (Wassarman *et al.*, 2001).

To assess sRNA expression under relevant environmental conditions, RNAs from five stages of growth in standard LB media from exponential to stationary phase and from two conditions known to strongly induce the expression of the major SPI-1 (Lee & Falkow, 1990; Song *et al.*, 2004) or SPI-2 (Deiwick *et al.*, 1999) virulence regions were probed. The results of this analysis are summarized in Table 5.2 (the whole set of candidates tested is shown in Table 10.8 in the Appendix). Including the 26 previously detected *Salmonella* sRNAs (Altier *et al.*, 2000a; Figueroa-Bossi *et al.*, 2006; Fortune *et al.*, 2006; Padalon-Brauch *et al.*, 2008; Papenfort *et al.*, 2006, 2008; Pfeiffer *et al.*, 2007; Sharma *et al.*, 2007; Viegas *et al.*, 2007), a total of 67 *Salmonella* sRNAs can now be considered to be experimentally validated.

It was determined whether eight of the new *Salmonella* sRNAs showed an Hfq-dependent pattern of transcript abundance that correlated with Hfq binding (Fig. 5.6 A). The STnc290, 440, 490, 520, 540 and 560 sRNAs were all enriched by Hfq coIP (Table. 5.2) by factors up to 51-fold (STnc440). The expression of the four sRNAs with the highest enrichment factors (STnc290, 440, 520, 560) was strongly reduced in Δ*hfq* and so classified as Hfq-dependent; in contrast, the accumulation of STnc150, STnc490 and STnc540 ($\geq$=1.0-, 5.1-, and 3.3-fold enrichment, respectively) was unaffected in the absence of Hfq. Three sRNAs expressing stable transcripts of $\approx$85 to 90 nts originate from close to, or within, IS200 transposable elements (Fig. 5.6 B). STnc290 and STnc440 are expressed just upstream of *tnpA_4* and *tnpA_6*, respectively, whereas STnc490 is antisense to the translational start site of the IS200 transposase ORF. IS200 elements generally posses two stem-loop structures, one of which is a Rho-independent transcription terminator that prevents read-through from genes located upstream of the integration site (Beuzón *et al.*, 1999). Given their location, the STnc290 sRNA could originate from processing of the STM3033 transcripts reading into the *tnpA_4* terminator structure; by analogy, STnc440 would be derived from STM4310 transcripts. If so, this would constitute interesting cases in which transposon insertion has created stable sRNAs. The other IS200 stem-loop functions as a translational repressor by sequestering the start codon of the transposon ORF (Beuzón *et al.*, 1999); STnc490 overlaps with this structure on the opposite strand and by acting as an antisense RNA may function as an additional repressor of IS200.

Furthermore, on Northern blots 10 of the 31 newly identified *Salmonella* sRNAs could be detected under the environmental conditions that were tested (Tables 5.2 and 10.8). These sRNAs yielded stable transcripts, predominantly in the 50 to 100 nt range. Generally, with one exception only candidates represented by $\geq$20 cDNAs in a cDNA pool yielded a signal on Northern blots (Tables 5.2 and 10.8). While this suggests some correlation between intracellular abundance and cDNA frequency, the case of STnc150 was observed, for which a single cDNA was recovered yet a strong signal was obtained on Northern blots. In contrast, several candidates with $>$20 cDNAs failed the Northern blot validation (Table 10.8). The corresponding cDNAs were probably derived from 5' or 3' UTRs of larger mRNA transcripts. This was tested on Northern blots of agarose gels (data not shown). 14 of such candidates had the appropriate orientation to flanking mRNA genes for being UTR-derived. Six of these showed signals ranging in size from 500 to 2000 nucleotides (STnc180, Stnc190, STnc330, STnc470, STnc610, and STnc640; Table 10.8 in the Appendix) and are likely to be processed mRNA species.

In addition to the sRNAs listed above, the cDNAs included two loci predicted to encode small peptides, *i. e.* shorter than the 34 amino acid cut-off used to define ORFs in the current *Salmonella* genome annotation (McClelland *et al.*, 2001). These are referred to as STnc250 and STnc570 in Ta-

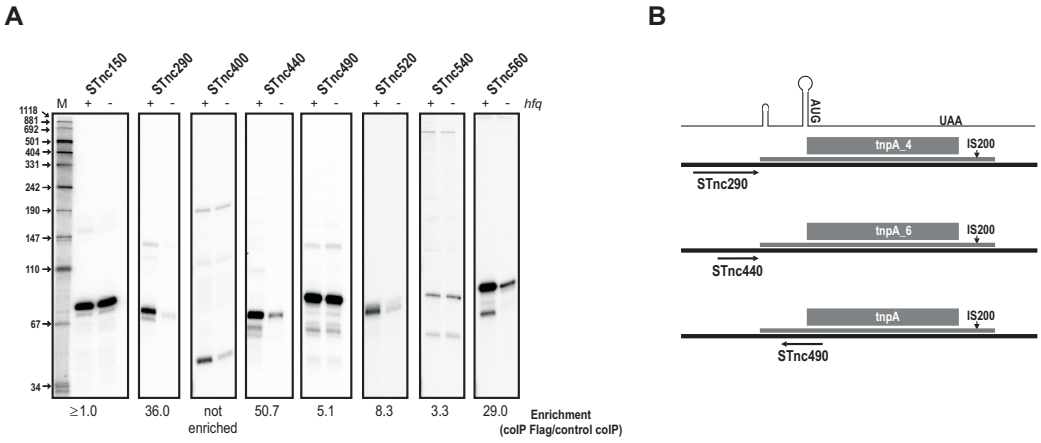


Figure 5.6: Expression of novel sRNAs in *Salmonella*. (A) RNA abundance of selected new sRNAs in wild-type (+) versus *hfq* mutant (-) *Salmonella* cells at early stationary phase (OD_{600} of 2). The enrichment factor of each of these sRNAs in the coIP experiment is given below the blots for comparison. **(B)** A schematic presentation of the position of the sRNAs according to the IS200 element is shown to the right. The upper drawing indicates the two stem-loop structures, start codon, and stop codon of the transposase-encoding mRNA of the IS200 elements. The three detected sRNAs are indicated by black arrows. (Figure was adapted from Sittka *et al.*, 2008.)

ble 5.2 and correspond to the predicted small *ypfM* and *yneM* mRNA-encoding genes, respectively, of *E. coli* (Wassarman *et al.*, 2001).

5.2. Discussion

To understand how bacterial RNA binding proteins such as Hfq mediate the control of global gene expression at the post-transcriptional level, direct targets need to be identified. These targets include, besides direct mRNA targets, the small noncoding RNA regulators which mediate post-transcriptional regulation of diverse mRNAs. In this Chapter, deep sequencing was applied to identify *Salmonella* RNA ligands bound to Hfq. This approach not only recovered most of the known sRNAs but also identified novel sRNAs and mRNA targets. Identification of novel *Salmonella*-specific sRNAs is particularly interesting as they could have a potential role in virulence regulation.

The first global approach for the identifcation of Hfq-bound RNAs involved detection of RNA co-immunoprecipitated with Hfq-specific antibodies on high-density oligonucleotide microarrays and identified new *E. coli* sRNAs (Zhang *et al.*, 2003). Similarly, microarray-based detection following co-immunoprecipitation of eukaryotic mRNA-protein complexes (mRNPs) identified endogenous organization patterns of mRNAs and cellular proteins (Tenenbaum *et al.*, 2002). Epitope-tagging of the yeast La homologue was successfully used for global coIP analysis (Inada & Guthrie, 2004). However, the requirement of custom high-density microarrays and/or species-specific antibodies has impeded similar studies in other organisms. The ideal sRNA discovery approach would not only detect sRNAs but would also define their exact sequence. Given the typical genome size of model bacteria ($\approx$5 Mb), a high-density oligonucleotide microarray with $\approx$10 million oligonucleotide

probes would be required to achieve single basepair resolution. Such arrays do not exist for any organism, and even today's high-density arrays (with 0.5 million features) come with extraordinarily high set-up and printing costs and are available for very few bacteria. The deep sequencing strategy remedies these technical and financial limitations.

The identification of Hfq-associated RNAs in *Salmonella* is based on a chromosomal epitope-tagging approach (Uzzau *et al.*, 2001), followed by coIP with a commercially-available antibody, and sequencing of hundreds of thousand cDNAs. The earlier shotgun-cloning studies in bacteria (see Section 2.3.5) and many other organisms (reviewed in Hüttenhofer, 2005; Hüttenhofer *et al.*, 2002) were limited by costly Sanger-type sequencing of individual cDNA inserts from plasmid vectors. The deep sequencing approach described here avoids a cloning step and is able to detect small RNAs with unparalleled sensitivity by defining the 5' and 3' ends of transcripts at basepair resolution.

Deep sequencing of cDNAs has identified the small RNA component of eukaryotic transcriptomes (*e. g.*, Lu *et al.*, 2005; Ruby *et al.*, 2006) and new classes of noncoding RNAs associated with eukaryotic RNA-binding proteins (Aravin *et al.*, 2006; Girard *et al.*, 2006). Furthermore, immunoprecipitation and deep sequencing has been successfully applied to characterize small RNAs associated with human Ago1 and Ago2 proteins, which are part of the eukaryotic RNA-induced-silencing-complex (RISC), and also to identify bound mRNA targets (Beitzinger *et al.*, 2007). In addition, an unexpected class of RNAs that originate from small nucleolar RNAs (snoRNAs) was identified in Ago complexes (Ender *et al.*, 2008). Thus, coIP followed by deep sequencing allowed to identify novel RNA binding partners. These studies in eukaryotes primarily focussed on the class of 20-30 nucleotide long microRNAs and short interfereing RNAs and typically included size-fractionation steps. Bacterial riboregulators are considerably larger (80-280 nucleotides), and in this study it was shown that even without prior size fractionation deep sequencing can capture and define the termini of these large sRNAs.

Analysis of deep sequencing results of Hfq-bound RNAs in *Salmonella* extends the tally of confidently identified sRNAs to 67 in this model pathogen (Table 5.2). Thirty eight of these are conserved sRNAs that were initially identified in *E. coli*, but only a few of their homologues have previously been shown to be expressed in other enteric bacteria (Altier *et al.*, 2000b; Bossi & Figueroa-Bossi, 2007; Fortune *et al.*, 2006; Julio *et al.*, 2000; Papenfort *et al.*, 2006, 2008; Sharma *et al.*, 2007; Viegas *et al.*, 2007). The finding, from this and other studies, that highly-conserved sRNAs are commonly expressed should prove useful to researchers working in other bacterial systems. A significant number of the Hfq-associated cDNAs correspond to sRNA loci that are absent from *E. coli* (Padalon-Brauch *et al.*, 2008; Pfeiffer *et al.*, 2007 and Table 5.2). Of these, *invR* exemplifies a sRNA gene that was probably horizontally acquired with the SPI-1 virulence region, early in *Salmonella* evolution (Pfeiffer *et al.*, 2007). Intriguingly, InvR is the most frequently recovered sRNA (>3,000 cDNAs in the Hfq-coIP library), which shows that this approach is not only effective in detecting conserved but also species-specific sRNAs of acquired pathogenicity regions. Horizontal transfer of virulence islands is a driving force in the evolution of bacterial pathogens (Dobrindt *et al.*, 2004), and knowledge of the functional elements of these islands is key to understanding pathogenesis. Whereas ORF identification in such islands has become routine, island-specific sRNAs are more difficult to recognize by bioinformatic-based approaches. These RNA factors could interconnect expression of the *Salmonella* core genome and virulence regions at the post-transcriptional level.

Besides confirming InvR, the present study found evidence for the expression of five of the 47 *Salmonella* sRNA candidate loci listed by Pfeiffer *et al.* (2007) who predicted orphan promoter/terminator pairs in IGRs (Table 10.8). Padalon-Brauch *et al.* (2008) recently reported the discovery of 18 *Salmonella* expressed sRNA loci. cDNAs of eight of these sRNAs could be recovered (IsrB-1, C, E, I-L, and P; Table 5.2). The fact that ten of these sRNAs were not recovered probably reflects their low-level expression under the growth condition used here (Padalon-Brauch *et al.*, 2008). This observation suggests an improvement that could be made to this method. RNomics- or microarray-based sRNA discovery methods require sRNAs to be expressed under the chosen assay condition, unlike bioinformatics-aided approaches that score for orphan transcription signals and primary sequence conservation (Argaman *et al.*, 2001; Chen *et al.*, 2002; Livny *et al.*, 2006; Wassarman *et al.*, 2001) or for conservation of RNA structure (Rivas *et al.*, 2001). However, as most sRNAs are involved in specific stress responses, it remains challenging to identify the appropriate growth condition that induces expression of certain sRNAs. Thus, future studies combining several different growth conditions with increasing sequencing depth are likely to identify even more novel sRNAs.

Similar to other global methods for RNA identification (Altuvia, 2007; Hüttenhofer & Vogel, 2006), this approach is likely to show certain biases regarding its specificity. For example, unspecific Hfq binding partners can be caused by cross-hybridization in the immunoprecipitation step or from the limited ability of reverse transcriptase to deal with stable RNA structures in cDNA synthesis, and these will need to be studied in more detail. However, it is clear that deep sequencing resolved the termini of many expressed and/or Hfq-bound sRNAs at basepair resolution (Fig. 5.5), which has not been achieved by other methods.

In addition to the Hfq-bound sRNAs, also the $\approx$35,000 reads derived from Hfq-bound mRNA targets ($\approx$19% of all reads in the Hfq-coIP; see Table 5.1 and Fig. 5.3) were analysed in the complete study which was published in Sittka *et al.*, 2008. This led to the identification of 727 Hfq-bound mRNAs. For a lot of mRNAs, especially sequences in the 5' UTRs or in intergenic regions of polycistronic mRNAs were specifically enriched in the Hfq-coIP libraries and often corresponded to known sRNA binding sites. Furthermore, the deep sequencing approach for the detection of Hfq-bound mRNAs was validated by comparison with the conventional approach, namely, hybridization of the RNA from the coIP samples to a *S. typhimurium* oligonucleotide microarray. Nearly half (45%) of the 365 enriched mRNAs in this coIP-on-Chip experiment correspond to regions identified by the deep sequencing approach. The overlap increased to 67% when genes that showed enrichment values above 5 were taken into consideration. Although coIP-on-Chip displays a lower sensitivity than deep sequencing these two independent methods do generate comparable results for the identification of mRNA-protein interactions.

In the complete study (Sittka *et al.*, 2008), the 454 sequencing results were additionally compared to mRNA profiles of the *Salmonella* wild-type and the *hfq* deletion strain using transcriptomic analysis on whole-genome microarrays. In total, expression of at least 785 genes, or 18% of the *Salmonella* genome, was changed either directly or indirectly by Hfq, and 32% of the affected mRNAs were bound to Hfq in the deep sequencing analysis. Conversely, 33% of the Hfq-bound mRNAs showed an Hfq-dependent pattern of gene expression. The proportion of genes identified to be Hfq-dependent is similar to *Pseudomonas aeruginosa* ($\approx$15% of all genes; Sonnleitner *et al.*, 2006) but bigger than for *E. coli* (6.3%; Guisbert *et al.*, 2007) or *Vibrio cholerae* (5.6%; Ding

et al., 2004). However, the different growth conditions and scoring parameters used for these other organisms preclude a direct comparison with the data obtained for *Salmonella* in this work. With regard to functional analysis of several Hfq-bound mRNAs, the combined transcriptomic and coIP data revealed that Hfq exerts a direct role in gene expression through the control of specific checkpoints in well-defined transcriptional regulons.

Transcriptomic profiling by itself is clearly unable to differentiate between transcriptional and post-transcriptional effects of Hfq. However, enrichment of a regulated mRNA in the Hfq-coIP library can successfully hint at post-transcriptional regulation by sRNAs. The current data set comprises several hundreds of such candidate mRNAs (see Table S4 in Sittka & Vogel, 2008) including many experimentally confirmed targets of *Salmonella* sRNAs. Integrating the score for Hfq-association deduced from present experiments and, where applicable, from the available *E. coli* data (Zhang *et al.*, 2003) into available algorithms such as TargetRNA (Tjaden *et al.*, 2006) could significantly improve target predictions for the large class of Hfq-dependent sRNAs.

A recent study based on sample-matched transcriptomics and proteomics by Ansong *et al.* (2009) also found that $>$20% of all annotated *Salmonella* genes are regulated post-transcriptionally either directly or indirectly by Hfq or SmpB, an RNA-binding protein which specifically interacts with tmRNA (Karzai *et al.*, 1999). For example, based on their proteomics analysis, expression of 781 proteins was found to be affected by Hfq, of which 25% overlap with the Hfq-associated mRNAs that were identified in this deep sequencing analysis.

Collectively, the method presented provides a first picture of the impact of Hfq on *Salmonella* gene expression at both the transcriptional and post-transcriptional level. A more detailed inspection of this data set, in particular of the $\approx$60% of the chromosome that remain to be fully analysed as well as sampling under different growth conditions, will probably expand the gamut of *Salmonella* small mRNA and noncoding RNA genes. In addition, the data sets could help to discover whether Hfq controls the expression of *cis*-antisense sRNAs that overlap with mRNA coding regions (Kawano *et al.*, 2005a) or whether certain *Salmonella* tRNAs are selectively associated with this protein (Lee & Feig, 2008; Scheibe *et al.*, 2007). Furthermore, the same approach was now also successfully applied to analyse RNA-binding properties of foreign Hfq proteins expressed in *Salmonella* (Sittka *et al.*, 2009). This *in vivo* approach identified endogenous *Salmonella* sRNAs as a major target of the foreign Hfq proteins from two eubacteria (*Neisseria meningitides, Aquifex aeolicus*) and an archaeon (*Methanocaldococcus jannaschii*) and additionally identified further novel sRNAs from *Salmonella*. In addition, specific RNA processing defects, *e. g.*, suppression of precursor processing of SraH sRNA, or aberrant accumulation of target mRNAs of the *Salmonella* GcvB, MicA or RybB sRNAs were observed. This indicated that the expression of heterologous RNA-binding proteins combined with deep sequencing analysis of RNA ligands can be used as a molecular tool to dissect individual steps of RNA metabolism *in vivo*.

Bacterial genomes encode a large number of RNA binding proteins (Anantharaman *et al.*, 2002) including globally acting proteins such as the CsrA/RsmA (Babitzke & Romeo, 2007) and Csp families (Yamanaka *et al.*, 1998). The combination of epitope-tagging and high-throughput sequencing of immunoprecipitated RNA could be used to identify the RNA targets of these proteins in any genetically tractable bacterium. This could help to unravel the post-transcriptional regulons in a wide range of bacteria.

CHAPTER 6

DEEP SEQUENCING REVEALS THE PRIMARY TRANSCRIPTOME OF *Helicobacter pylori*

The intense study of *Helicobacter pylori* (see Section 2.6.2), one of the most prevalent human pathogens, has contributed much to understanding bacterial virulence mechanisms. The availability of the $\approx$1.67 Mb *H. pylori* genome sequence has greatly facilitated these studies, including the discovery of proteins with important functions in gastric infections. In comparison, much less is known about the overall transcriptional organization and the noncoding regions of the *H. pylori* genome. Genome annotation of *H. pylori* strain 26695 predicted 1,590 open reading frames (Tomb *et al.*, 1997), yet experimental studies over the years identified only $\approx$55 promoters in this organism (see Table 10.11 in the Appendix). Thus, it remains largely unknown where and when the transcription of the majority of the *H. pylori* genes is initiated. In addition to the lack of clear transription signals, *Helicobacter* contains only a very limited repertoire of transcriptional regulators, *e. g.*, only three sigma factors, three two-component systems, and two additional orphan response regulators (Section 2.6.2). However, in order to cope with the various stresses it encounters during infection, *e. g.*, pH and nutrient fluctuations, *H. pylori* must have additional mechanisms to regulate its gene expression.

Besides protein regulators, small noncoding RNAs (sRNAs) have been identified in all kingdoms of life as modulators of gene expression, and several systematic screens have led to the identification of sRNAs in diverse bacteria (see Section 2.3). So far, no small regulatory RNAs have been described in the ϵ-subdivision of proteobacteria, except for the highly conserved housekeeping RNAs, tmRNA, RNaseP RNA, and SRP RNA (see Figure 6.1). Homology searches based on sRNA sequences from other bacteria have not been helpful in identifying sRNA candidates in *H. pylori*, suggesting that either sRNA genes are absent or have diverged beyond the limit of detection. Due to a lack of a known Hfq homologue (Sun *et al.*, 2002), the approach based on combination of epitope-tagging and HTPS of immunoprecipitated RNA that was described for identification of novel sRNAs in *Salmonella* in the previous Chapter could not be applied to *Helicobacter*. This indicated a requirement of *de novo* identification approaches, such as bioinformatics-based predictions (Section 2.3.3) or direct RNomics (Section 2.3.5).

Previous RNomics screens for sRNA identification in bacteria, *e. g.*, *E. coli*, the two archaeons *Sulfolobus solfataricus* and *Archaeoglobus fulgidus*, and *Aquifex aeolicus*, were at that time based on the conventional Sanger sequencing method (Section 2.3.5). The development of deep sequencing

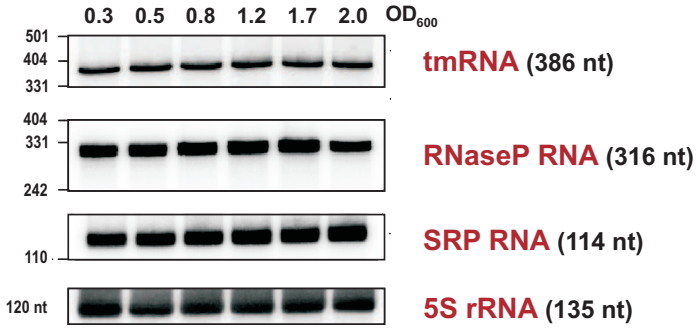


Figure 6.1: Northern Blot of *Helicobacter* housekeeping RNAs. Total RNA was extracted from *H. pylori* 26695 cultures in different growth phases and probed with DNA oligos (listed in Table 10.14) on Northern blots. Samples were taken at OD_{600} values between 0.3 (log phase) and 2.0 (late stationary phase). Equal amounts of RNA were loaded. The *Helicobacter* homologues of tmRNA, RNase P RNA, SRP RNA, and 5S rRNA were initially predicted by Brown (1999); Regalia *et al.* (2002); Williams & Bartel (1996), and Tomb *et al.* (1997), respectively.

technologies allows a more complete and rapid coverage of small RNA profiles in diverse organisms, *e. g.*, in plants using massively parallel signature sequencing (Lu *et al.*, 2005). Furthermore, high-throughput pyrosequencing (Margulies *et al.*, 2005) of size-fractionated RNA has been successfully applied to the discovery of small RNAs, *e. g.*, in *C. elegans* (Ruby *et al.*, 2006) and the single-cell algae *Chlamydomonas* (Molnár *et al.*, 2007; Zhao *et al.*, 2007). With the advent of next-generation sequencing technologies it is now possible to generate tens of millions of short sequences in a single assay. This development has enabled the recent 'RNA sequencing' (RNA-Seq) technology via random cDNA libraries, which was successfully applied to the transcriptomes of *Arabidopsis* (Lister *et al.*, 2008), mouse embryonic stem cells (Cloonan *et al.*, 2008) and other mouse tissues (Sultan *et al.*, 2008), budding (Nagalakshmi *et al.*, 2008) and fission yeast (Wilhelm *et al.*, 2008) at single nucleotide resolution.

In this study, an initial biocomputational approach based on prediction of orphan promoter/terminator pairs in intergenic regions (IGRs) was used for the identification of sRNAs in *H. pylori*. However, it led to the identification of only a small number of sRNAs in *H. pylori* and, hence, a direct RNomics approach was chosen. To specifically enrich cDNA libraries for primary transcripts and by depleting processed RNAs, such as highly abundant rRNA fragments, a method based on an exonuclease which digests RNAs with a 5'-monophosphate was developed. Subsequently, the *H. pylori* transcriptome was analysed by sequencing a total of ≈ 3.7 million cDNAs derived from *H. pylori* grown under standard laboratory and stress conditions or in contact with eukaryotic cells. This identified many additional sRNAs both from intergenic regions and from regions antisense to annotated open reading frames (ORFs) as well as the ubiquitous 6S RNA and its associated pRNAs. Moreover, differential analysis of primary and processed RNA species facilitated the identification of ≈ 800 transcription start sites of mRNAs across the *H. pylori* genome. The results of the *H. pylori* transcriptome analyses will improve the functional annotation of the *H. pylori* and related genomes, and the approach used here should facilitate the global transcriptome analysis of mixed pathogen-host populations.

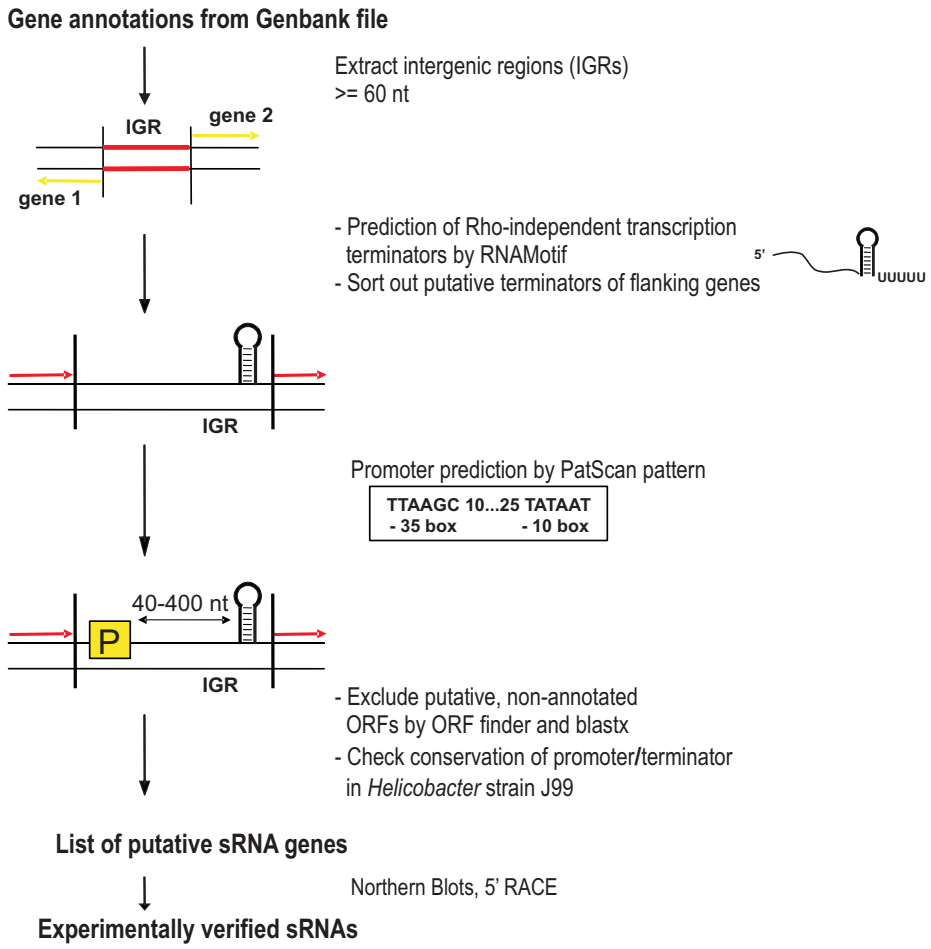


Figure 6.2: *In silico* prediction of novel sRNAs in *H. pylori* 26695. To predict novel sRNA candidates in *H. pylori* 26695, intergenic regions ≥ 60 bp were scanned for orphan promoter/terminator pairs. Rho-independent transcription terminators were predicted using `RNAMotif` (Lesnik *et al.*, 2001) and the descriptor file for the terminator model specified in Chen *et al.* (2002). Putative promoters were determined by simple pattern searches for *H. pylori* promoter motifs with `PatScan` (Dsouza *et al.*, 1997). After removal of putative, non-annotated ORFs, sRNA candidates that were conserved in the *H. pylori* strain J99 were subjected to experimental confirmation by 5' RACE and Northern Blots.

6.1. Results

6.1.1. Biocomputational prediction of ncRNAs in *H. pylori*

As a first approach to identify novel sRNA candidates in *Helicobacter pylori* strain 26695, a bioinformatics-based strategy similar to Chen *et al.* (2002) and Argaman *et al.* (2001) was undertaken (see Section 2.3.3). Specifically, intergenic regions ≥ 60 bp were scanned for orphan promoter/ terminator pairs as outlined in Figure 6.2. Since there is only little information about promoter consensus sequences and transcription termination signals in *Helicobacter*, this approach was based on similarities to known sRNA features from *E. coli*. Most of the known *E. coli* sRNAs are

terminated by Rho-independent transcription terminators, which is a common termination mechanism also for mRNAs in this bacterium. However, two previous studies predicted only few Rho-independent terminators in *H. pylori* (Ermolaeva *et al.*, 2000; Washio *et al.*, 1998). In contrast, it has been recently shown that *Helicobacter* might rely more on Rho-independent termination than previously assumed (de Hoon *et al.*, 2005; Kingsford *et al.*, 2007).

For this first screen, Rho-independent transcription terminators were predicted using the `RNAMotif` descriptor file specified by Lesnik *et al.* (2001) that was also previously used for sRNA prediction in *E. coli* (Chen *et al.*, 2002). Additionally, putative promoter sequences were identified by simple pattern searches for *H. pylori* RpoD (σ^{80}) promoter motifs using `PatScan` (Dsouza *et al.*, 1997; Vanet *et al.*, 2000). Only promoter and terminator pairs within a distance of 40-400 nt were considered as sRNA candidates. After removal of putative, non-annotated ORFs and conservation analysis of candidates in *Helicobacter pylori* strain J99 this rather constrained screen ended up with a final list of six candidates (Intergenic regions (IG): IG75, IG433, IG449, IG480, IG494, and IG550) for experimental verification by 5' RACE and Northern blot analyses (see Table 10.10 in the Appendix).

The 5' ends of four candidates (IG75, IG433, IG480, and IG550) could be confirmed by 5' RACE (indicated in the alignments in Figures 10.7-10.11 and for the *H. pylori* 26695 specific IG550 candidate in Fig. 10.12 in the Appendix). However, only one candidate, IG480, gave a strong signal on Northern Blots (see Figure 6.3A), while for IG75, 433 and 550 only weak signals could be detected (data not shown). The identified sRNA IG480 is also included in the list of sRNAs predicted in *Helicobacter* by Livny *et al.* (2006). `BLAST`[1] searches revealed that this RNA has five additional homologues in the *H. pylori* 26695 genome (see Figure 6.3 B, asRNA B to F), whereof four could be detected on Northern Blots under the conditions examined (Fig. 6.3 A). These additional homologues were missed in the bioinformatics-based predictions because they either overlap annotations and are, hence, not included in the list of IGRs or did not pass the terminator threshold.

As two of the identified homologues (asRNA E and F) are located antisense to ORFs annotated as short hypothetical proteins (HP0024 and HP1515), the identified sRNAs could act as antisense RNAs and thereby inhibit expression of the peptides. Thus, they were named *H. pylori* **a**nti**s**ense RNA (asRNA) A to F. Alignment of the antisense RNAs revealed the presence of highly conserved Rho-independent terminators and promoter elements that are in accordance with transcriptional start sites mapped by 5' RACE (Fig. 6.4 A & B). Otherwise, the sequences are less conserved. However, all of the sRNAs are predicted to fold into two highly stable stem loop structures (Fig. 6.4 C).

Alignment of the nucleotide sequences of the opposite strand of the IGRs carrying the antisense RNAs A, C, D, E, and F revealed a highly conserved Shine-Dalgarno sequence followed by an AUG start codon (Fig. 6.5 A). The corresponding RNAs encode all for short peptides of 30 amico acids and were, thus, termed **s**hort **p**eptide RNAs (spRNAs). Only spRNA D encodes for a shorter peptide of 23 aa, as a deletion of one nucleotide in the coding sequence (position 176 in the alignment in Fig. 6.5 A) leads to a premature stop codon and a truncated peptide. The spRNAs C and D could be detected *in vivo* as $\approx$ 230 nt long transcripts on Northern blots (see Fig. 6.5B), and also expression of spRNA A and F was confirmed on Northern Blots (data not shown). The peptides encoded by spRNA E and F correspond to the ORFs annotated as HP0024 and HP1515 in the *H. pylori*

[1] `blast.ncbi.nlm.nih.gov/Blast.cgi`

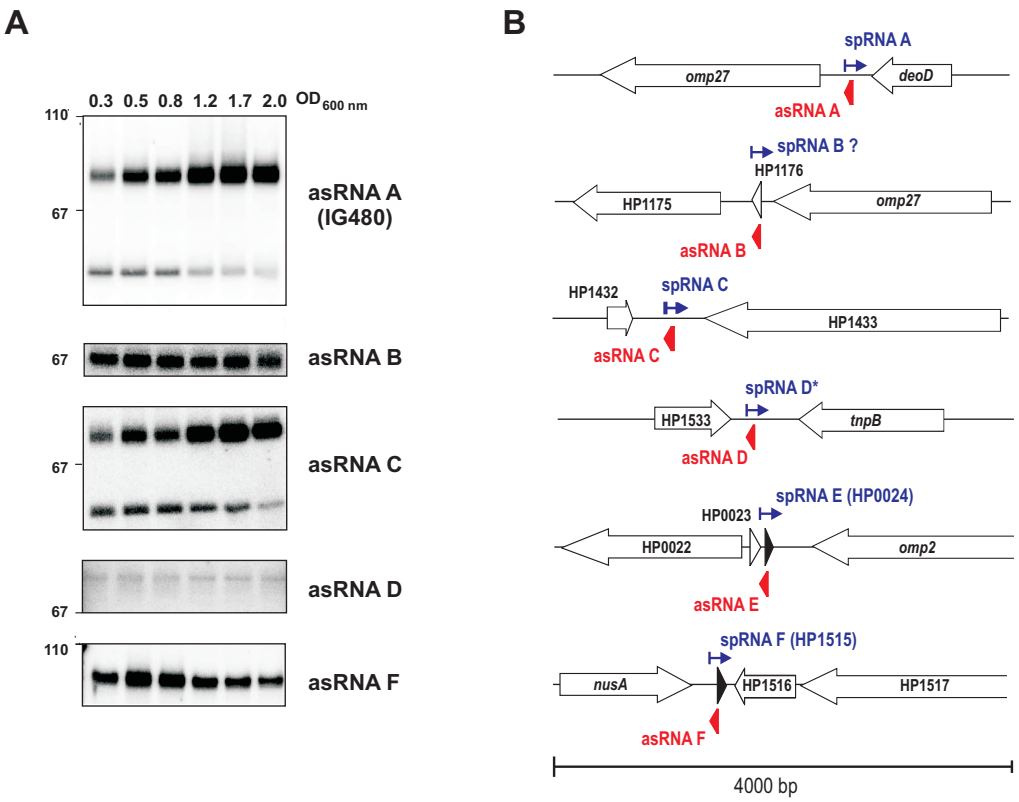


Figure 6.3: A family of small RNAs in *H. pylori* 26695. (A) Total RNA samples of *H. pylori* strain 26695 from OD_{600} 0.3 to 2 were separated on denaturing PAA gels and analysed on Northern blots. Under these conditions, signals have been observed for asRNAs A to F except for asRNA E. The homologues, A through F, vary in length between ≈ 67 and ≈ 90 nt. Bands corresponding to smaller species in the blots for sRNA A and C are likely to be processing products. DNA oligonucleotides used for Northern blot hybridization are listed in Table 10.14. **(B)** Location of asRNA A to F in the *H. pylori* 26695 genome. *H. pylori* asRNA B overlaps with an ORF encoding a short hypothetical protein (HP1176), whereas asRNA E and F are located antisense to ORFs annotated as short hypothetical proteins (HP0024 and HP1515). Blue arrows indicate short peptide encoding RNAs (spRNAs) A to F.

genome but the other three ORFs have not been annotated yet. As the two annotated ORFs are indicated as hypothetical proteins and also BLAST searches did not reveal any similarity to other proteins, the function of these peptides is unclear. However, the peptides A, C, E , and F contain a lot of positively charged amino acids, lysine (K), arginine (R), or histidine (H), leading to a positive net charge between '+6' and '+8'. Moreover, as the peptides are very hydrophobic and transmembrane domains could be predicted using the TMPred[2] program (shaded in grey for each protein in Fig. 6.5C), an interaction with membranes is very likely. In addition, the peptides can probably form alpha helical structures and were predicted to be potential antimicrobial peptides using the prediction program on the Antimicrobial Peptide Database[3] (Wang & Wang, 2004). Thus,

[2] www.ch.embnet.org/software/TMPRED_form.html

[3] aps.unmc.edu/AP/main.php

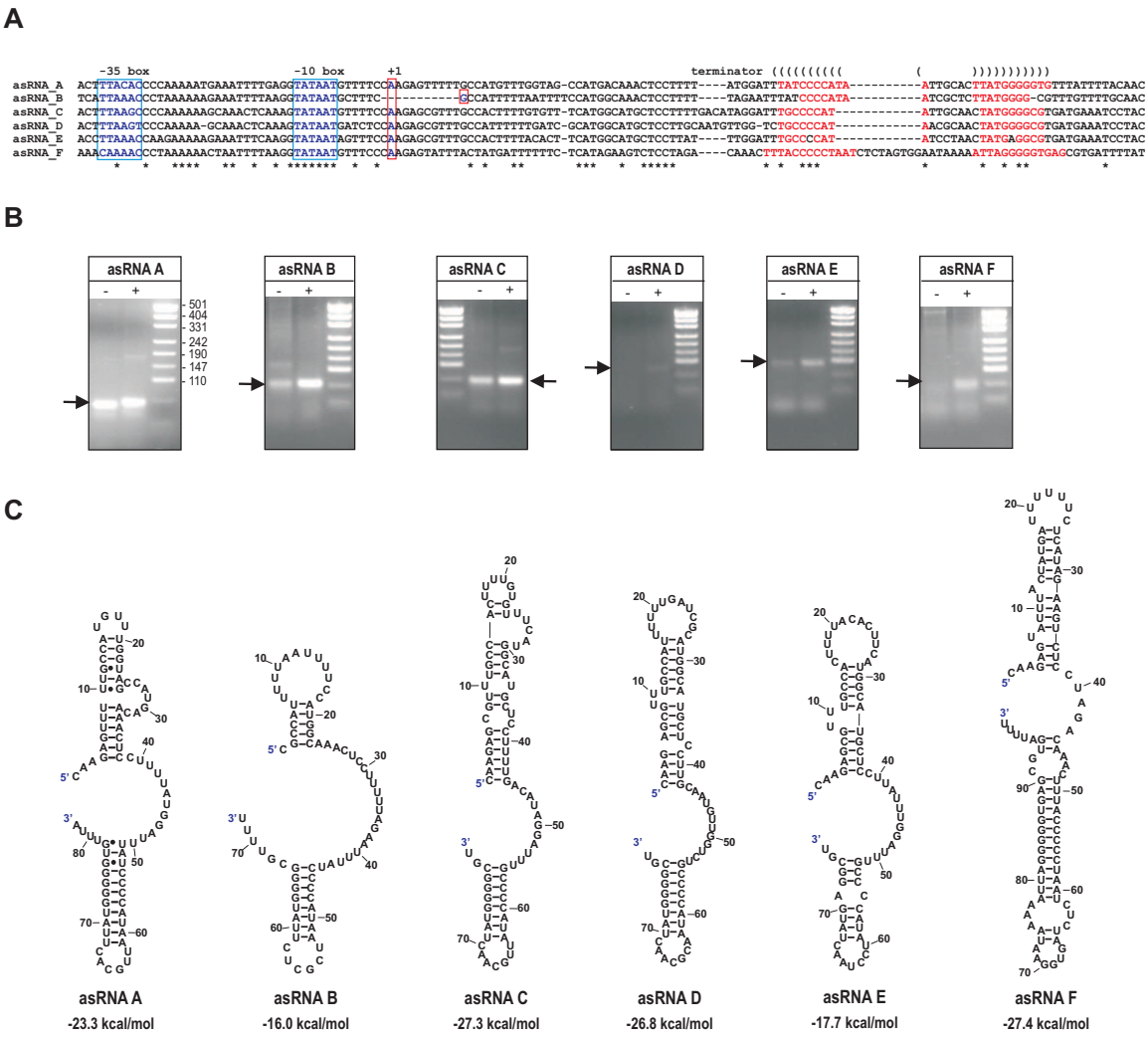


Figure 6.4: Sequences and structures of a family of small RNAs in *H. pylori* 26695. (A) Alignment of the family of sRNAs shows high conservation of promoter elements and predicted Rho-independent terminators. Brackets indicate paired bases within the terminator stem-loop structure of asRNA A; red letters indicate terminator base-pairs in each sequence. Promoter elements are indicated by blue boxes, transcriptional start-sites determined by 5' RACE in **B** by red boxes, respectively. Asterisks indicate conserved nucleotides. **(B)** The predicted 5' ends were verified by 5' RACE analysis. **(C)** All of the asRNAs are predicted to fold into two highly stable stem loop structures. Secondary structures were computed with `RNAstructure` (Mathews *et al.*, 2004) and drawn with `RnaViz` (Rijk *et al.*, 2003).

they could eventuelly act as toxins for *H. pylori* itself or, in case they are secreted, also be toxic for other bacteria in the mucosa.

Conservation analysis by `BLAST` searches showed that all of the asRNAs/spRNA cassettes are present in several *Helicobacter* strains (see Figure 6.6) and some of them, asRNA C and F, are present in several copies in *H. pylori* J99. For some of the spRNAs, mutations within the ORF lead to truncated peptides or a start codon mutation results in a complete loss of translation. For example, spRNA D which is truncated in *H. pylori* 26695 is not truncated in the other strains, whereas for

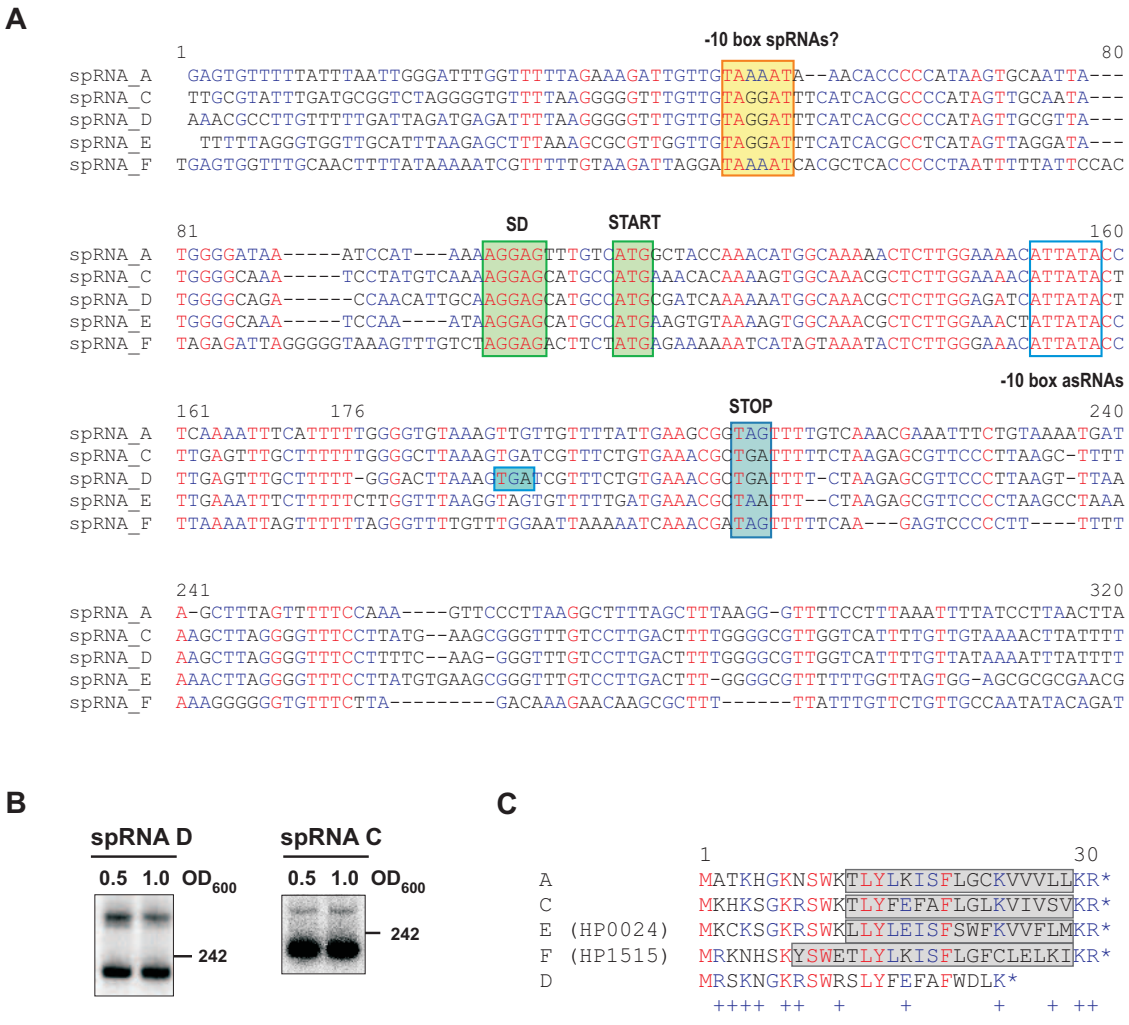


Figure 6.5: A family of small ORFs in *H. pylori* 26695. (A) Nucleotide alignment of a family of small ORFs shows high conservation of the Shine-Dalgarno sequence and start codon (green boxes) as well as the stop codon (light blue box). The -10 promoter element of the short peptide encoding RNAs is highlighted in yellow, the -10 box of the asRNAs encoded on the opposite strands is framed in blue. **(B)** Expression of spRNA C and D was confirmed on Northern blots. **(C)** Protein alignment of peptides A, C, D, and E. Transmembrane domains were predicted using the `TMPred` program and are shaded in grey for each protein. '+' indicates a positively charged amino acid (H, K, or R) in at least one sequence in the alignment.

example spRNA A harbours a STOP mutation in *H. acinonychis*. Thus, the peptides could have redundant functions and probably expression of not all peptides is required in *H. pylori*. Overall, the asRNAs are very likely to act as *cis*-encoded regulators that bind to the spRNAs and thereby probably repress translation of the peptides and/or lead to degradation by the double-strand specific RNase III.

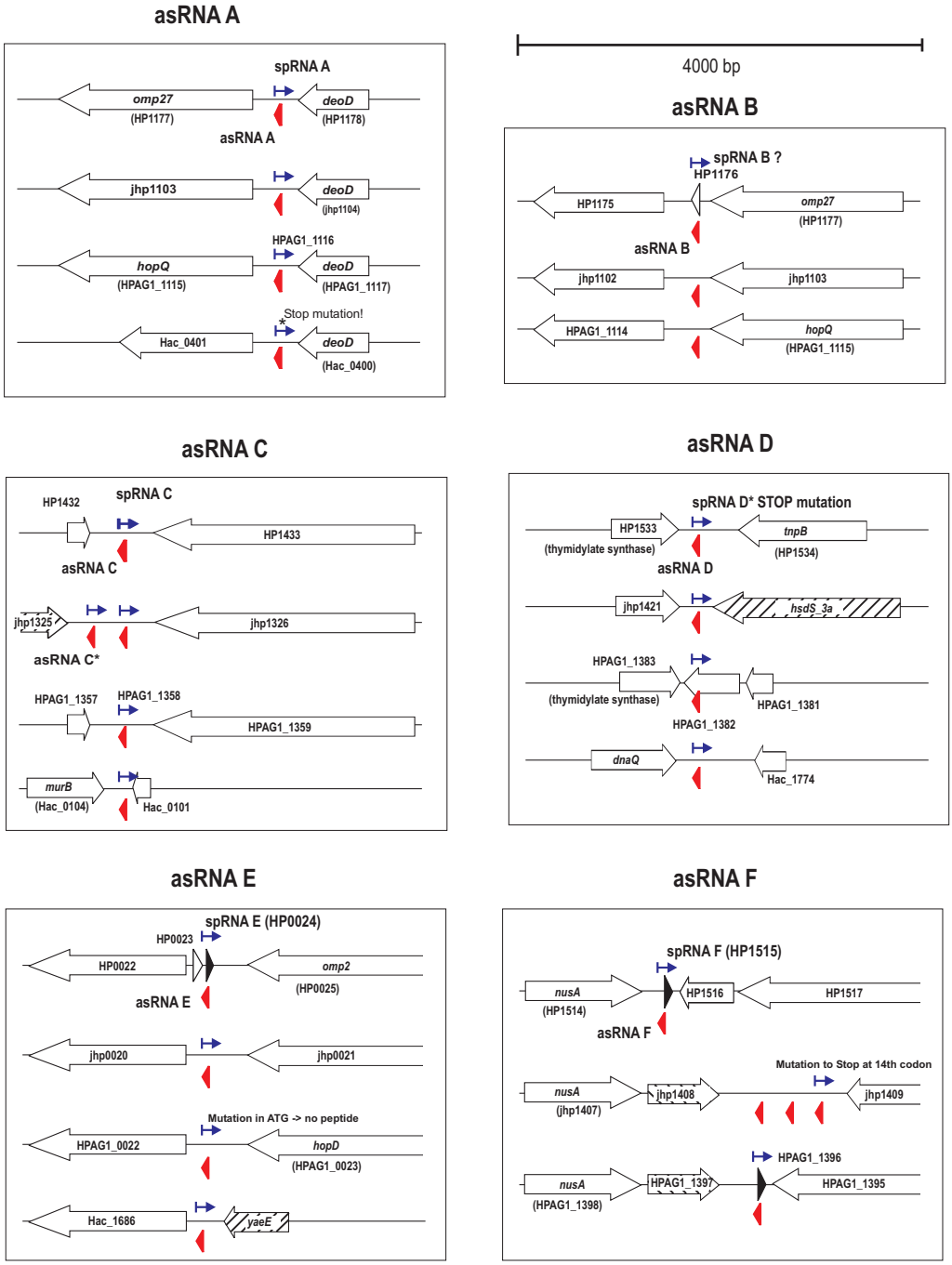


Figure 6.6: Conservation of antisense RNAs A to F and short ORF cassettes. Locations of homologues of asRNA A to F and peptide encoding RNAs spRNA A to F on the opposite strand in diverse *Helicobacter* strains. HP: *H. pylori* 26696, jhp: *H. pylori* J99, HPAG1: *H. pylori* HPAG1, and Hac: *Helicobacter acinonychis*. Stripes indicate inserted genes or different flanking genes than in strain *H. pylori* 26695.

6.1.2. Depletion of processed RNAs

The biocomputational strategy based on approaches for the identification of sRNAs in *E. coli* predicted only a few sRNAs in *Helicobacter*. However, sRNAs in *Helicobacter* might have different sequence and secondary structure features than sRNAs from *E. coli*, and a *de novo* approach to identify sRNAs might be required. Thus, a direct RNomics approach based on massively parallel pyrosequencing of cDNA libraries derived from a variety of growth conditions was chosen. Instead of size-fractionating RNA as it was done in previous RNomics screens in bacteria based on Sanger sequencing methods (Section 2.3.5), a method for depletion of processed RNAs and selective sequencing primary transcripts was developed.

In bacteria, newly initiated (primary) mRNA and sRNA transcripts carry a characteristic 5' triphosphate (5'PPP) end, whereas processed RNA species, including abundant ribosomal and transfer RNAs (rRNA, tRNA), have 5' monophosphate (5'P) ends. Treatment of total RNA with 5' monophosphate dependent terminator exonuclease (TEX) which degrades 5'P but no 5'PPP RNA depletes these processed RNAs and in parallel enriches for primary transcripts (Fig. 6.7 A).

To selectively identify primary transcripts in total RNA samples of *H. pylori*, two differential cDNA libraries for each of the tested conditions were constructed: one library (-) from the original RNA pool covering both primary and processed transcripts, and a second one (+) from RNA treated with TEX in which primary transcripts are enriched. After TEX treatment of total RNA samples of *Helicobacter* and *Salmonella*, the bands for 23S and 16S rRNA are almost completely eliminated (Fig. 6.7 B). Northern Blot analysis confirmed specific degradation of a 23S rRNA fragment in *Helicobacter* (Fig. 6.7 C). In contrast, 5S rRNA is neither in *H. pylori* nor in *S. typhimurium* accessible for TEX-mediated degradation (Fig. 6.7 D). Probably its 5' end is masked in a very stable, unaccessible secondary structure. However, primary transcripts, such as asRNA A, sRNA B (identified below, see Section 6.1.7), and the *Salmonella* sRNA InvR (Pfeiffer *et al.*, 2007) are not degraded. In contrast, a processed fragment of *Salmonella* SraH RNA, which was originally identified in several genome-wide screens in *E. coli* (Argaman *et al.*, 2001; Wassarman *et al.*, 2001), is completely degraded as expected (Fig. 6.7 D). The protection of asRNA A is not as prominent as for other sRNAs, *e. g.*, sRNA B or InvR. Probably, for part of the asRNA A transcripts, the 5'-triphosphate was cleaved by an endogenous pyrophosphatase or the first nucleotides were cleaved off by an endonuclease in both cases leading to a 5'-monophosphate. The processed fragment of *Helicobacter* asRNA A is not degraded, as it is derived from the 5'end and, thus, still carries the 5'-triphosphate. Based on these preliminary experiments, total RNA treated with TEX should be specifically enriched for sRNAs and, in addition, probably also for primary transcripts of mRNAs.

6.1.3. Deep sequencing of *Helicobacter* cDNA libraries

Helicobacter cDNA libraries were prepared from five different growth conditions: culture in brain-heart-infusion media to mid-log phase (C-/+ libraries), and following 30 min acid stress (AS-/+); growth in cell culture flasks in the absence (PL-/+) or presence of two eukaryotic cell types, *i. e.* AGS human gastric epithelia cells (AGS-/+) in which *H. pylori* induces motility and cellular elongation upon contact, and Huh7 cells (Huh7-/+) which the bacteria can adhere to but fail to induce morphological changes (see Figure 6.8). In total, 2.15 million reads were sequenced for the acid

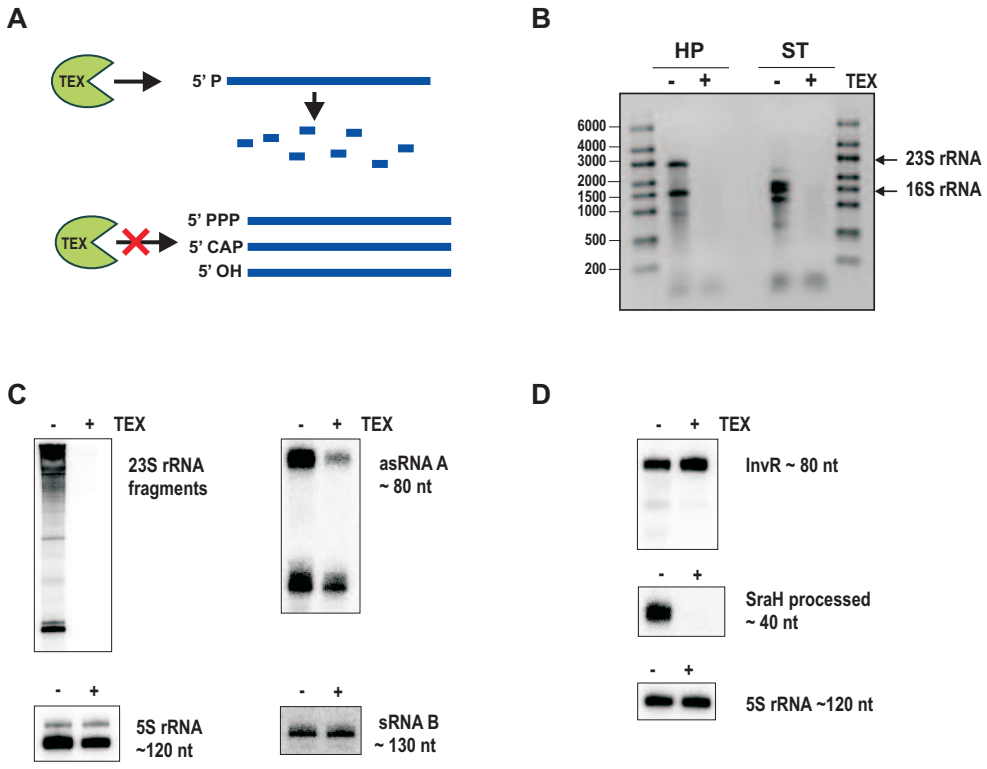


Figure 6.7: Enrichment of primary transcripts. (A) 5' monophosphate dependent terminator exonuclease (TEX) specifically degrades RNAs with 5'-monophosphates, while primary transcripts with a 5'triphosphate group are not affected. **(B)** Treatment of total RNA with TEX eliminates most of the processed RNAs, especially ribosomal RNAs (indicated by arrows). Total RNA from *H. pylori* harvested at an OD_{600} of 0.6 (HP) and total RNA from *S. typhimurium* harvested at an OD_{600} of 2.0 (ST), respectively, were separated on a TBE-agarose gel and stained with ethidium bromide; -/+ indicates prior treatment with TEX. **(C)** Northern blots for several *Helicobacter* RNAs -/+ TEX treatment. **(D)** Northern blots for several *Salmonella* RNAs -/+ TEX treatment. DNA oligonucleotides used for Northern Blot probing are listed in Table 10.14 in the Appendix.

stress libraries and 1.79 million reads for the infection libraries on a Roche FLX sequencer. This resulted in 220,000-530,000 cDNAs per library and a total of $\approx$3.7 million cDNAs (see Table 8.15 in Material and Methods).

After 5'end linker clipping, reads were mapped to the *Helicobacter* genome using the program `segemehl` which is based on an error-tolerant suffix array method (Hoffmann *et al.*, 2009, submitted). For this mapping method, clipping of tailing sequences is not necessary as they will be removed during the mapping step. However, for very short sequences the poly(A) tail often leads to mapping errors. Therefore, a filtering step was introduced which removed all sequences with an A-content of more than 70% (see Table 8.15 in Material and Methods). For these sequences, the poly(A) tail was clipped separately. Of the clipped reads, all sequences $\geq$12 nt were mapped again with `segemehl`. This procedure allowed to map also very short sequences of at least 12 nucleotides. In total, between 62% and 84% of the reads for the individual libraries could be mapped to the *Helicobacter* genome. Of the mapped reads, between 26% and 75% uniquely mapped to the

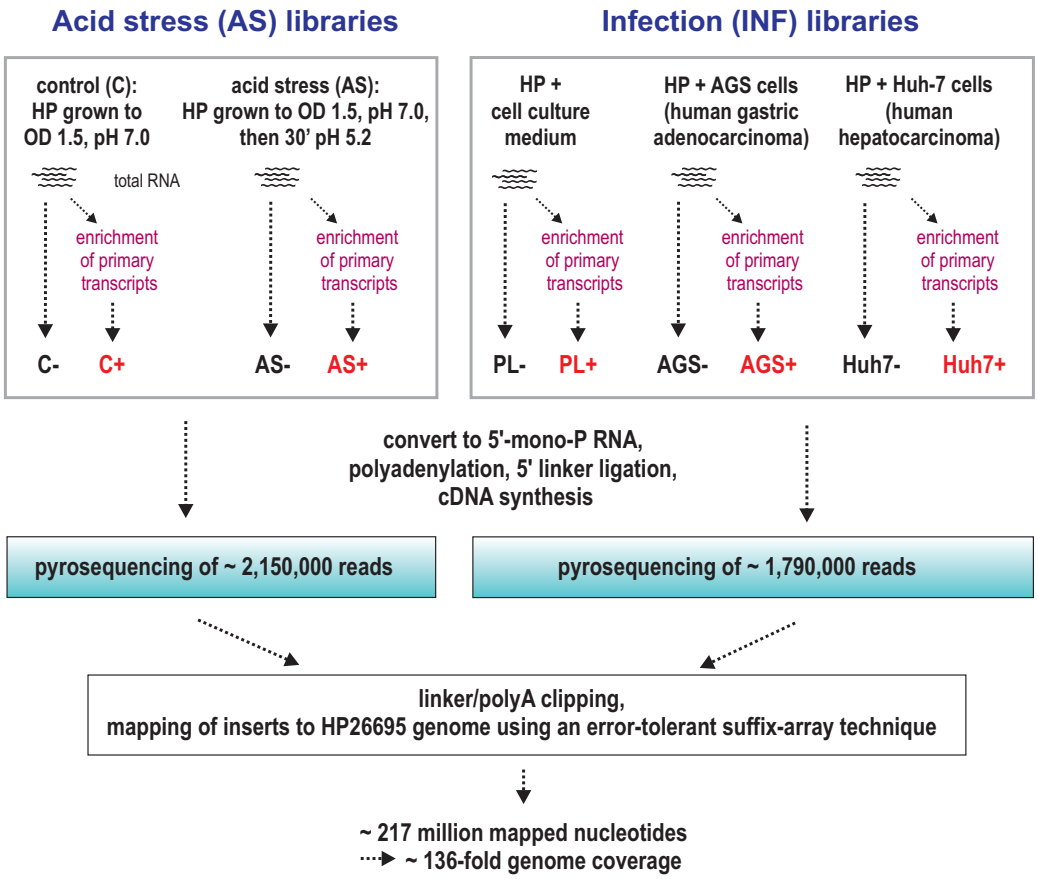


Figure 6.8: *H. pylori* 454 cDNA libraries. Total RNA of *H. pylori* from acid stress samples or infection samples was -/+ TEX treated. For cDNA library construction, RNA was converted to 5'-mono-P-RNAs by TAP (tobacco acid pyrophosphatas) treatment, followed by addition of poly(A)-tails, linker ligation and reverse transcription. Libraries were sequenced on a Roche FLX sequencer. After linker and poly(A)-tail clipping, reads were mapped to the *H. pylori* 26695 genome using an error-tolerant suffix array technique (Hoffmann *et al.*, 2009, submitted). In total, 217 million nucleotides were mapped which corresponds to a $\approx$136-fold genome coverage of the 1.67 Mb *H. pylori* genome.

Helicobacter genome (Table 8.15 in Material and Methods). The lengths of mapped reads varied in a range from 12 to $\approx$350 bp (Fig. 6.9 A and B), whereof 12 bp was the length cut-off for mapping.

Figure 6.10 shows for each library the distribution of reads $<$12 bp, reads with no match to the *H. pylori* genome, and reads that overlap with annotated regions (for actual numbers see Table 10.12 in the Appendix). The fraction of reads that could not be mapped to the *Helicobacter* genome is higher for the AGS-/+ and Huh7-/+ libraries as they also contain human reads from the host cells (see also Table 8.15 in Material and Methods). The fraction of uniquely mapped reads is generally higher in the enriched libraries due to the removal of a large number of ribosomal rRNA reads. Reads which derive from these transcripts map at least twice to the *Helicobacter* genome due to the presence of two 16S and 23S rRNA genes and three 5S rRNA genes. Figure 6.10 shows that the fraction of ribosomal RNAs (orange) is reduced in the enriched (+) libraries compared to the (-)

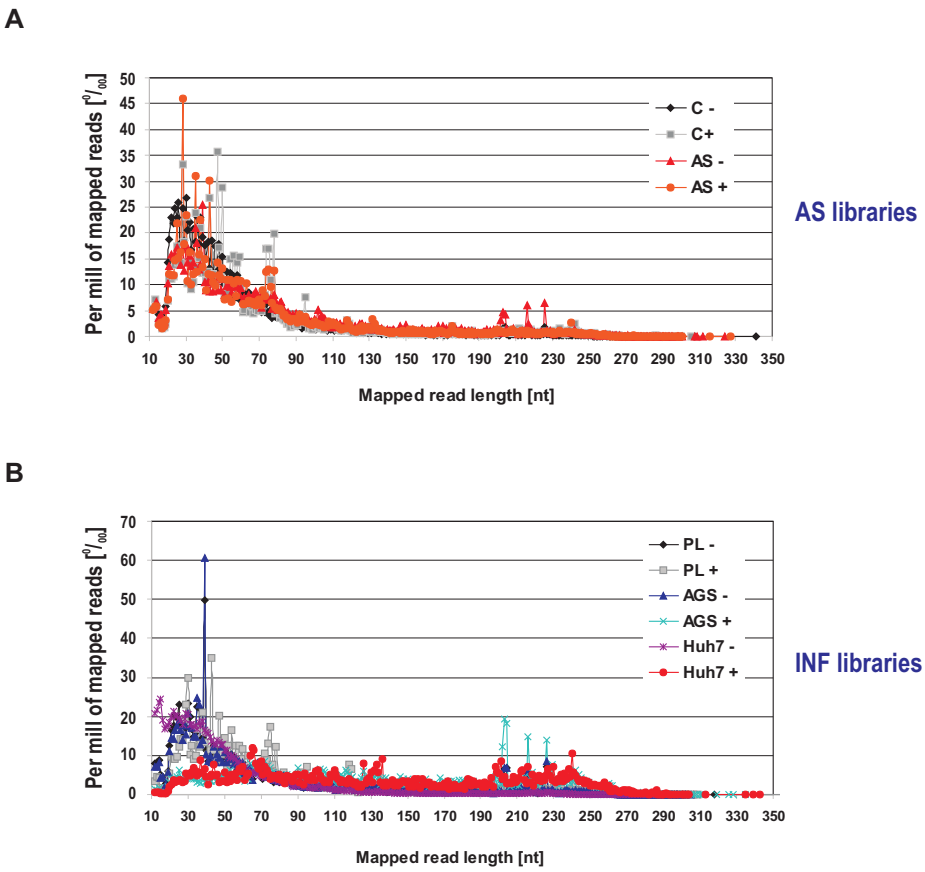


Figure 6.9: Mapped read length distribution of *H. pylori* 454 cDNA libraries. (A) and **(B)**: Mapped read length distribution for the acid stress (AS) and infection (INF) libraries, respectively. Per mill reads of each library are plotted *vs* the observed mapped read length.

libraries. For example, in the C- library ≈ 63% of all reads derive from rRNAs, but only ≈ 25% in the C+ library. In contrast, the fraction of tRNAs increased upon TEX treatment, *e. g.*, 7.3% tRNA reads in the C- and 27% tRNA reads in the C+ library, probably also due to a strong pairing of the 5' and 3' end within the tRNA structure. A large fraction of reads maps to mRNAs (8.8% and 5.8% for the C- and C+ libraries, respectively). This is most prominent in the acid stress libraries (32.5% and 18.9% for the AS- and AS+ libraries). Moreover, a fraction of reads mapping antisense to annotated genes was observed (5.4% and 7.9% for the AS- and AS+ libraries, respectively). These reads probably derive from *cis*-encoded antisense RNAs. In addition, the fraction of reads from intergenic regions contains promising candidates for novel sRNAs in *H. pylori* (*e. g.*, 3.4% and 15% for the C- and C+ libraries, respectively). Furthermore, the increase of reads in intergenic regions in the TEX+ libraries shows the successful enrichment of primary transcripts upon TEX treatment.

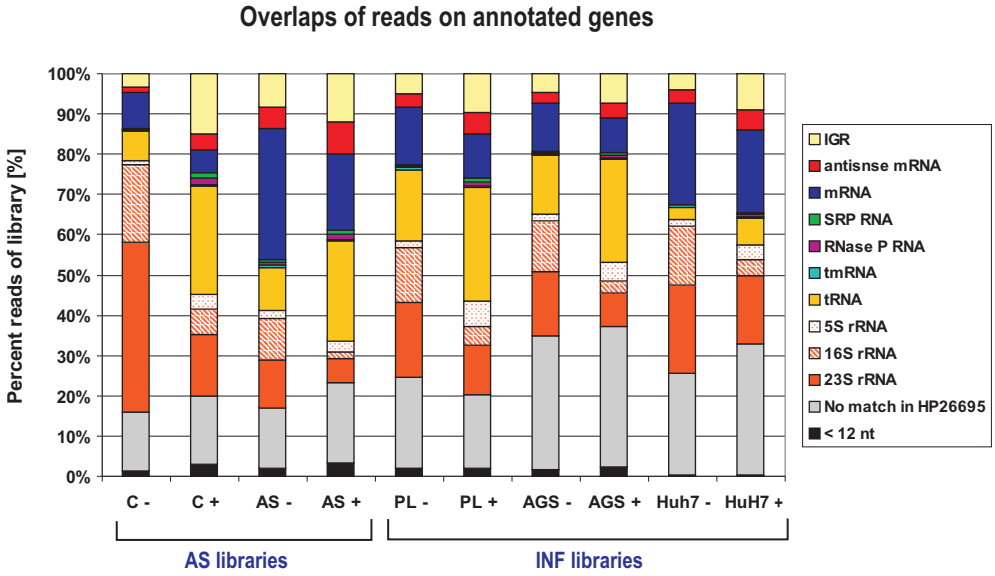


Figure 6.10: Read distribution on annotated genes. Bar diagrams showing the relative proportions of different RNA species in the acid stress (AS) and infection (INF) libraries.

Next, the number of cDNA hits for each nucleotide position in both genomic strands was calculated for all libraries similarly as described in Section 5.1.2 for the Hfq-CoIP libraries. The resulting graphs were visualized using the Integrated Genome Browser (IGB, Affymetrix). To get an overview of the read distribution along the *Helicobacter* chromosome, all positions mapped for the enriched libraries (+TEX) and all positions mapped for the untreated libraries (-TEX) were pooled and plotted along the plus and minus strand of the *H. pylori* 26695 genome (Figure 6.11A, upper panel). The distribution of reads along the *H. pylori* chromosome indicated an overall genome coverage.

The lower panel in Figure 6.11A shows the read distributions for the C-/+ and AS-/+ libraries on part of the *cag* pathogenicity island. While reads are distributed along the whole mRNAs in the unenriched libraries (black graphs), the 5'end of mRNAs are enriched in the TEX treated libraries. These enrichment peaks correspond exactly to experimentally mapped promoters for *cagA* and *cagB* as indicated by green arrows (Spohn *et al.*, 1997). Also the two known promoters for the urease operon, which has an important function in acid resistance of *Helicobacter*, show the sharp enrichment peaks in the TEX treated libraries compared to the untreated libraries (Fig. 6.11B). Transcription of this operon starts upstream of *ureA* (Shirai *et al.*, 1999; Spohn & Scarlato, 1999a) and upstream of *ureF* (Akada *et al.*, 2000; Pflock *et al.*, 2005). Furthermore, a strong induction of this operon in the acid stress libraries is visible. The lower amount of reads mapping to the 3' part of the *ureIEF*-HP0068-*ureH* polycistronic mRNA and the relatively sharp decrease of mapped reads within *ureE* in the untreated libraries indicate a previously described processing site in *ureE* (Akada *et al.*, 2000).

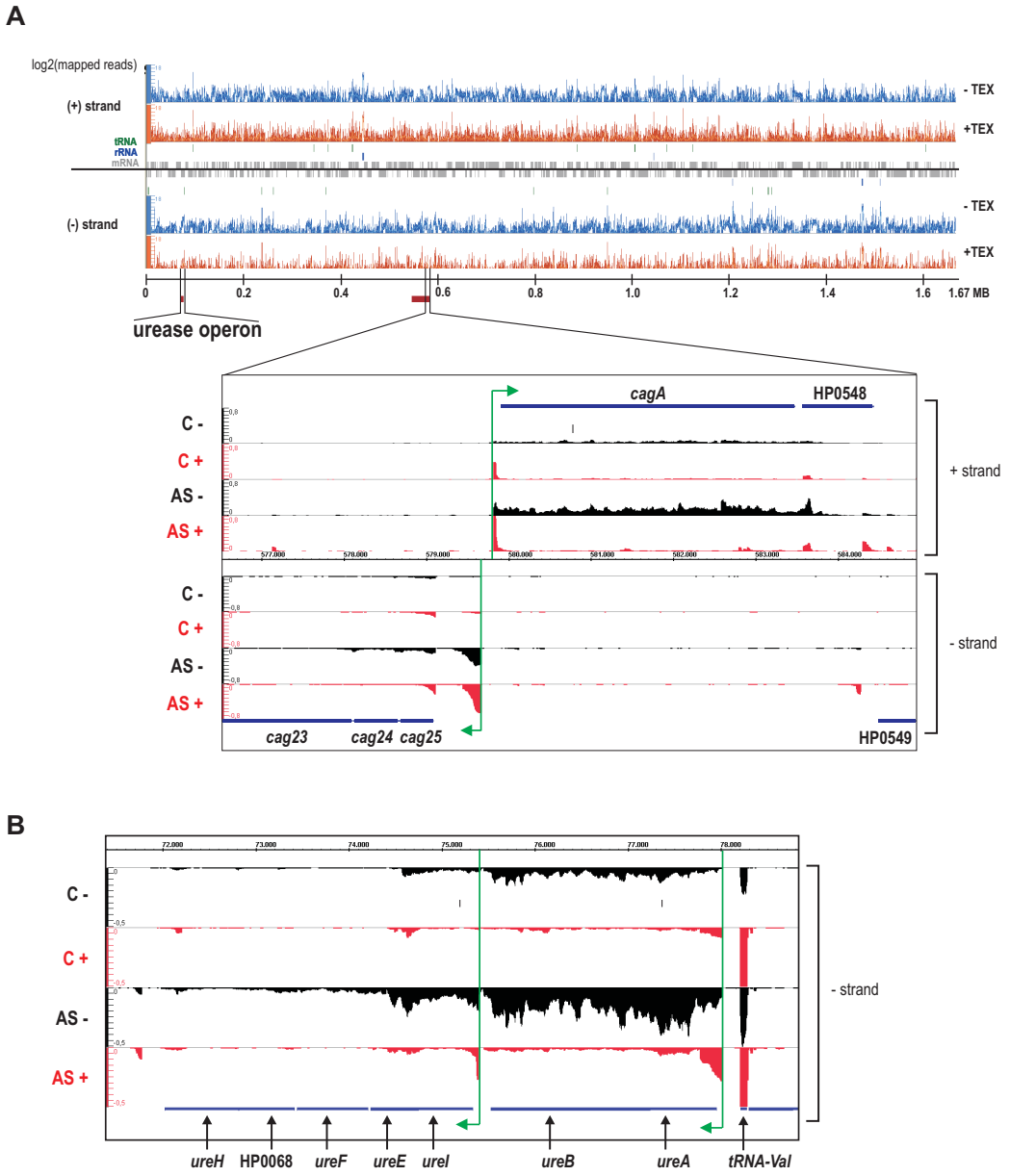


Figure 6.11: Cag island and urease operon. (A) *(Upper panel)* Screen shot of the IGB showing the read distribution for all untreated libraries (-TEX, blue curves) and all libraries enriched for primary transcripts (+TEX, orange curves). The reads for all untreated and for all enriched libraries were pooled, and the $\log_2$ values of the number of mapped reads for the plus and minus strand at each nucleotide position were plotted along the genome. Positions of tRNA, rRNA, and mRNA genes are indicated by green, blue, and grey bars, respectively. *(Lower panel)* Read distribution for the C-/+ libraries and AS-/+ libraries on part of the *cag* pathogenicity island. Green arrows indicate the experimentally mapped transcriptional start sites for *cagA/B* (Spohn *et al.*, 1997). Note that *cagB* is not annotated in *H. pylori* 26695 but located between *cag25* and *cagA* in strain G27. **(B)** Read distribution on the minus strand for the C-/+ libraries and AS-/+ libraries on the urease operon. Green arrows indicate the experimentally mapped transcriptional start sites for *ureA* (Shirai *et al.*, 1999; Spohn & Scarlato, 1999a) and *ureI* (Akada *et al.*, 2000; Pflock *et al.*, 2005).

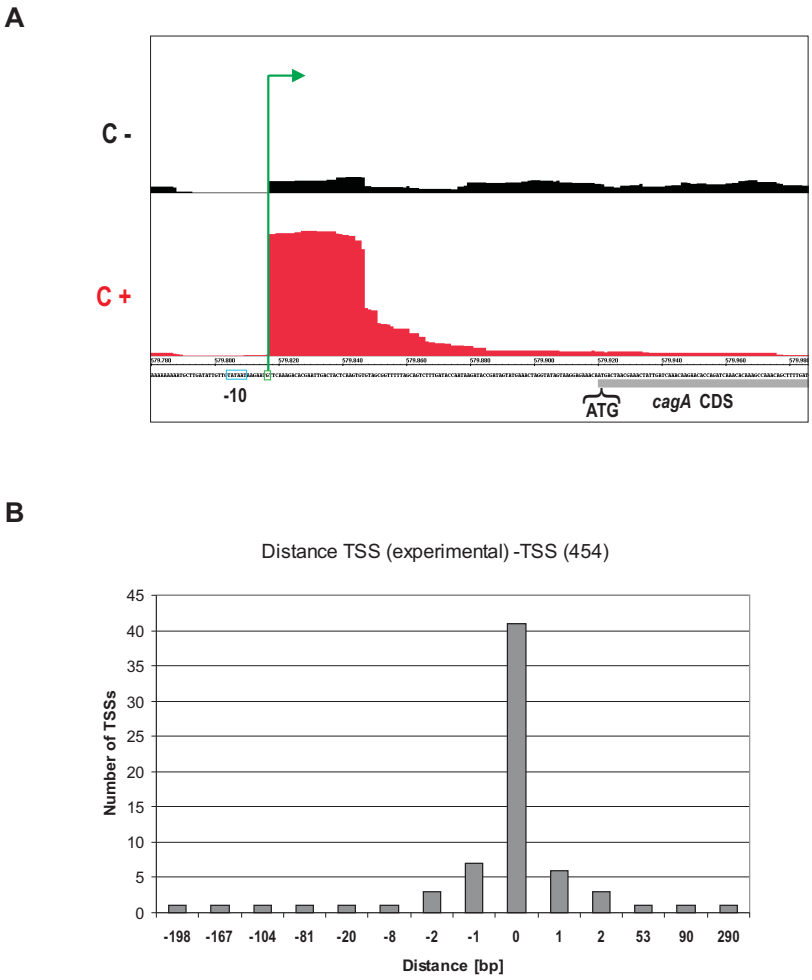


Figure 6.12: Enrichment patterns allow an exact mapping of transcriptional start sites. (A) A typical TSS enrichment pattern is show exemplarily for *cagA*. A strong enrichment of cDNAs clustering at the +1 site of *cagA* mRNA in the TEX-treated C+ (red) library could be observed, whereas in the untreated C-(black) library the reads are spread along the whole mRNA. The cDNA reads of the C+ library cluster exactly at the transcriptional start site published in Spohn *et al.* (1997), which is indicated by a green arrow. **(B)** In total, 74 experimentally mapped promoters from the literature or based on 5' RACE analysis (see Table 10.11 in the Appendix) were compared to the TSSs based on the 454 sequencing data. The histogram indicates the number of observed distances between experimental TSS and 454 TSS.

6.1.4. Mapping of transcriptional start sites

Figure 6.11 exemplifies how the different cDNA library data reveal both the transcriptional start site (TSS) and the growth-condition dependent regulation of a given gene. Mapping of the individual cDNA sequences to the *H. pylori* genome showed that nuclease-treatment had counter-selected processed rRNAs in favor of transcripts matching to intergenic regions. These latter regions also contain the TSSs of mRNA genes. Figure 6.12A shows the typical enrichment pattern of cDNAs clustering at the +1 site of an mRNA, here *cagA* mRNA. Intriguingly, the 5' ends of these cDNA

clusters matched exactly the previously mapped first nucleotide of *cagA* mRNA. In contrast, cDNAs from the TEX- library were dispersed over the 5' mRNA region. In other words, scoring for differences in the two cDNA libraries revealed a pattern indicative of the genomic TSS positions of primary transcripts (Figure 6.12A).

To assess the quality of TSS determination based on the 454 mapping, a set of 54 promoters was compiled from the literature and 20 additional promoters were mapped by 5' RACE (see Table 10.11 in the Appendix). Of these 74 TSSs, 69 were covered by the 454 data. Comparison between the experimentally determined start sites and the ones observed in the deep sequencing data shows that most start sites map exactly to the same nucleotide (Figure 6.12, distance of 0). For some genes, large differences were observed which probably correspond to alternative promoters. However, the significant overlap between experimentally verified TSSs from the literature based on primer extension or 5' RACE and TSSs from the 454 data gives good confidence to annotate new promoters based on the 454 data.

6.1.5. Global TSS annotation

To annotate TSSs in *H. pylori* at a genome-wide level, all libraries were simultaneously loaded into the IGB and altogether analysed by manual inspection. Generally, regions showing the typical enrichment pattern of the 5' flanking nucleotide in the (+) over the corresponding (-) library in at least two of the five samples were annotated as TSSs. For some genes no clear enrichment pattern was observed. In these ambiguous cases, additional criteria such as the position of a putative TSS to adjacent genes were included. For example, a TSS was annotated although it did not show a convincing enrichment pattern if it was located between two divergent genes and no other TSS was annotated for the respective gene. As the two genes are divergent, each must have its own promoter.

The global TSS analysis predicted a total of 1,906 putative TSSs, whereof 938 ($\approx$ 49%) were annotated on the minus strand and 968 ($\approx$ 51%) on the plus strand. As depicted in Figure 6.13A, all annotated TSSs were classifed into one out of five categories: primary (P), secondary (S), internal (I), antisense (A) or orphan (O). TSSs with the highest expression that are located $\leq$ 500 bp upstream on the same strand of an annotated gene were classified as primary. If a gene had an additional alternative transcriptional start site with lower expression, this TSS was classified as secondary. A TSS which started within an annotation on the same strand was classified as internal, whereas TSSs that were located on the opposite strand of an annotation or started on the opposite strand within a range of -/+100 bp of an annotation were classified as antisense. TSSs which had no related annotation in close proximity (-/+100 basepairs to annotations of the opposite strand or $>$500 basepairs to downstream genes on the same strand) were called orphan (Fig. 6.13A).

As outlined in Figure 6.13B, TSS can have multiple associations with different annotations, *e. g.*, the TSS within gene 3 is internal to gene 3 but also primary to gene 4. Thus, in total 2,495 annotated TSS features (1,265 on the (-) strand and 1,230 on the (+) strand, respectively) were annotated. The distribution of TSSs into the different classes is shown in Figure 6.13B. A large fraction of antisense transcripts was observed for the *Helicobacter* transcriptome. Around 27% of the primary transcripts (216 out of 810 unique primary TSSs) were also annotated as antisense to another gene (Fig. 6.13C). Conversely, $\approx$ 23% of the antisense TSSs are also primary TSSs (219 out of 969 unique antisense TSSs). Around 18% (145 out of 810) of the primary TSSs are also internal TSSs,

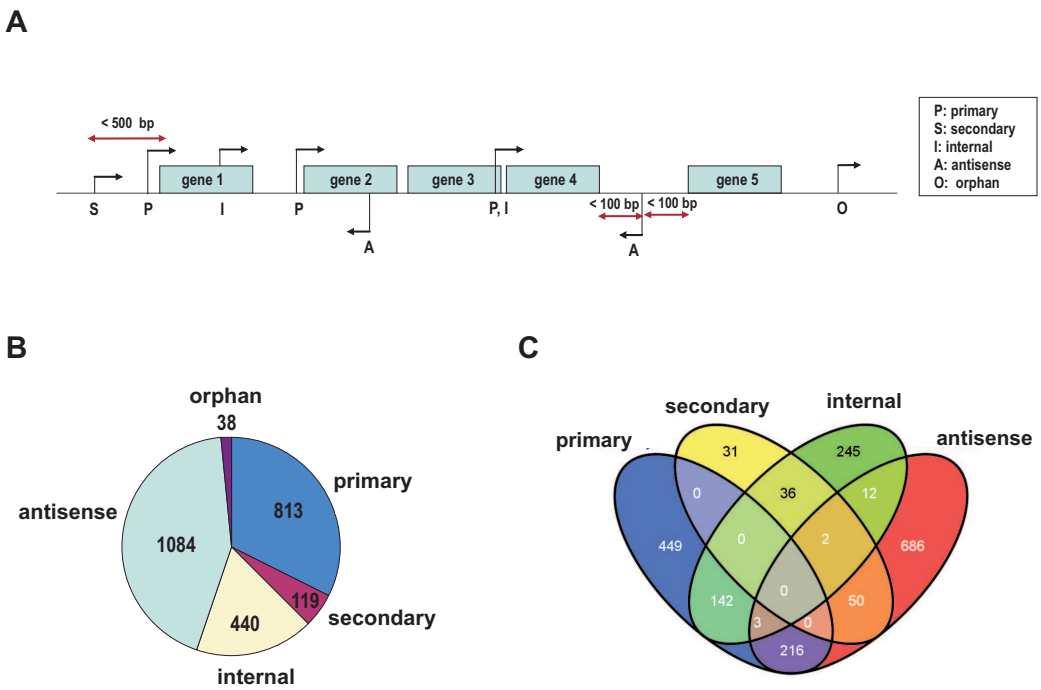


Figure 6.13: Global annotation of transcriptional start sites in *H. pylori*. **(A)** Schematic drawing of the criteria used for manual TSS annotation. All annotated TSSs were classifed into one of five categories: primary (P), secondary (S), internal (I), antisense (A) or orphan (O) based on their expression strength and distance to flanking or overlapping annotations. **(B)** Pie chart showing the relative proportions of different TSS classes. **(C)** Venn diagram showing the overlap between different TSS classes. Some TSSs can be associated to multiple genes and therefore be represented in different classes, *e. g.*, primary to gene x and antisense to gene y. The diagram was generated by VENNY (http://bioinfogp.cnb.csic.es/tools/venny/index.html) using the TSS positions for the four classes P, S, I, and A as input.

i. e. the TSS is located sense within an annotated gene, and conversely 33% (145 out of 440) of the internal TSSs are also primary. This observation indicates a very dense distribution of promoters along the *Helicobacter* genome and hints at the existence of alternative transcription start sites for genes that are located within operons.

For 717 of the 1,576 annotated *H. pylori* ORFs a primary TSS was annotated and 105 genes showed an additional secondary TSS. This set of 822 primary and secondary TSSs was used for further analysis, *e. g.*, calculation of the UTR length distribution (see Figure 6.14). Surprisingly, 25 of the 822 mRNAs ($\approx$ 3%) turned out to be leaderless, *i. e.* they lack a 5' UTR and their transcription starts exactly on the 'A' of the AUG start codon (Table 10.13 in the Appendix). Only a few mRNAs had a 5' UTR length $<$20 nt. A peak in the UTR length distribution was observed between 20 to 50 nt (396 TSS = 48%). This is in good accordance with the experimentally determined region known to be covered by the 30S subunit during translation initiation (Murakawa & Nierlich, 1989; Platt & Yanofsky, 1975; Steitz & Jakes, 1975). Upon anchoring to the Shine-Dalgarno sequence

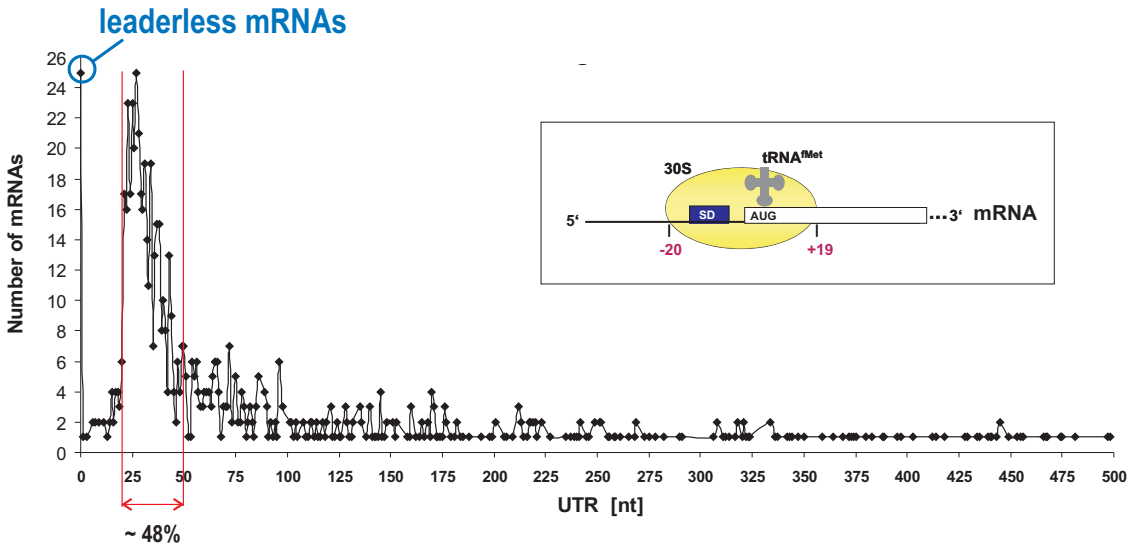


Figure 6.14: UTR length distribution. The distribution of UTR lengths was based on 822 primary and secondary TSSs of mRNA genes. The frequency of the observed UTR length is plotted *vs* the UTR length. In total, 25 leaderless mRNAs (5' UTR of 0 nt) were identified which are listed in Table 10.13 in the Appendix. Almost half of the 5' UTRs range in size between 20 and 50 nt. The inlet shows schematically the region of the mRNA that is covered during fomation of the translation initiation complex.

and the start codon, the 30S subunit covers a region ranging from ≈-20 bp upstream of the start codon to ≈+19 bp downstream of the start codon (Beyer *et al.*, 1994; Hüttenhofer & Noller, 1994; see inlet in Figure 6.14). More recently, X-ray analyses mapped the path of an mRNA through 70S ribosomes and indicated that a region of -/+15 nt around the start codon is wrapped around the 30S subunit and passes two separate tunnels (Yusupova *et al.*, 2006, 2001). Thus, for very short UTRs (<20 bp) the sequence upstream of the start codon is probably not long enough to contain a SD sequence or binding sites for the ribosomal protein S1. For several genes, long 5' UTRs with a length of up to 500 bp, which was the distance threshold for primary TSSs, were observed. These could harbour *cis*-regulatory elements such as riboswitches.

6.1.6. 454 sequencing reveals a 6S RNA homologue of *H. pylori*

One highly transcribed region in all libraries is located upstream of *purD* (HP1218) and antisense to the hypothetical protein HP1219 (Fig. 6.15 and 6.16A). From this region, an almost constitutively expressed RNA of ≈180 nt is transcribed (Fig. 6.16B). The hypothetical ORF HP1219 is not conserved in any of the other *H. pylori* strains and might therefore be misannotated (see Fig. 6.16A). The same region has recently been described to contain a candidate structured RNA motif identified by the `CMfinder` tool (Weinberg *et al.*, 2007). The authors suggested this RNA to be a riboswitch aptamer but could not detect any structural modulation by binding of a panel of purine compounds *in vitro*.

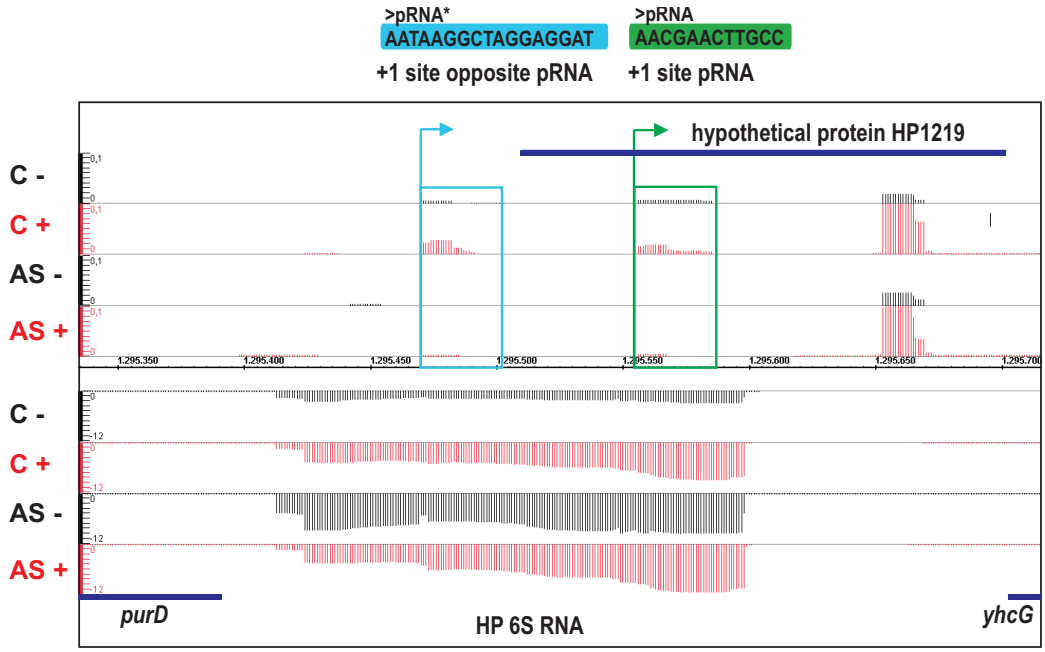


Figure 6.15: Deep sequencing reveals a 6S homologue in *H. pylori*. Screen shot of the IGB showing a highly transcribed region opposite to the hypothetical protein HP1219 in the C-/+ and AS-/+ libraries. The 180-nt long RNA upstream of *purD* is very likely a 6S homologue in *Helicobacter*. Furthermore, 6S-specific pRNAs on the opposite strand of 6S RNA are marked in green and additional pRNAs* are indicated in blue.

Secondary structure prediction and *in vitro* structure probing of the 180-nt long RNA, which is extended at the 5' end compared to the RNA motif identified by Weinberg *et al.*, 2007, revealed a long hairpin structure similar to 6S RNA from *E. coli* (Trotochaud & Wassarman, 2005) as shown in Figure 6.16C. The RNA polymerase inhibitor 6S RNA has been identified in almost every bacterial clade, but no 6S homologue has been described so far in the ϵ-subdivison of proteobacteria (Barrick *et al.*, 2005; Willkomm *et al.*, 2005). Although no sequence similarities to the *E. coli* 6S RNA can be found, the putative *H. pylori* 6S RNA contains some of the conserved secondary features, such as the large central bubble, a 1-nt bulge in the closing stem, and some of the conserved base-pairs of the previously described consensus structure (Barrick *et al.*, 2005). Indeed, consensus shape analysis (Reeder & Giegerich, 2005) of the regions upstream of the *purD* genes of other bacteria from the ϵ-subdivision, such as *Campylobacter*, *Wollinella*, as well as the deep-sea bacteria *Nitratiruptor* and *Sulfurovum*, indicated that they can also form into the 6S RNA like secondary structure (see Fig. 6.17). However, on the primary sequence level they show a high level of diversity compared to *Helicobacter*.

6S RNA adopts a structure similar to an open promoter complex and has recently been shown to be a template for RNA-mediated synthesis of pRNAs (Gildehaus *et al.*, 2007; Wassarman & Saecker, 2006). Thus, detection of these 14 to 20 nt-long pRNAs in the cDNA libraries would strongly argue for this RNA to be an 6S homologue. Indeed, several reads for pRNAs could be detected in the TEX treated library (Fig. 6.15). The *Helicobacter* pRNAs start exactly at the same position in the

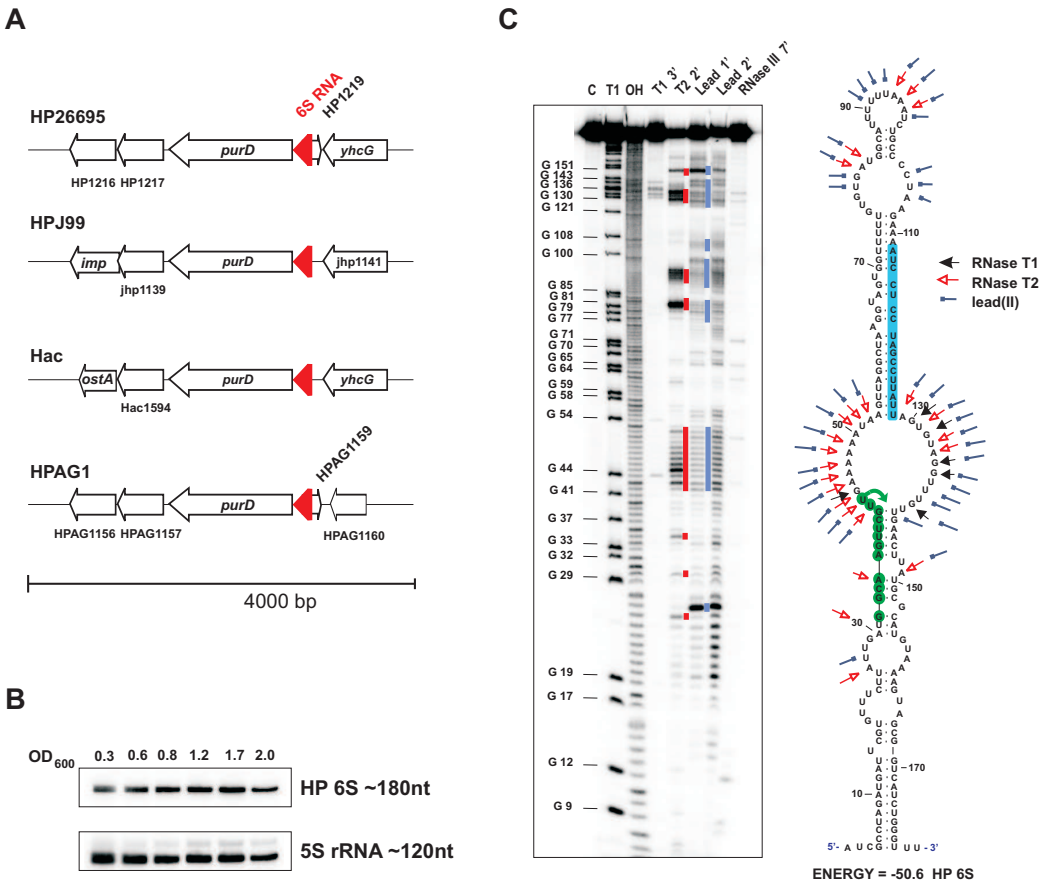


Figure 6.16: Genomic location, expression, and structure of *H. pylori* 6S RNA. (A) Genomic locations of putative 6S RNA homologues in diverse *Helicobacter* strains. The hypothetical protein HP1219 is not conserved in two other *H. pylori* strains and probably misannotated. **(B)** Expression of the 180-nt long 6S RNA was confirmed on Northern Blots. RNA samples were taken at OD_{600} values between 0.3 (log phase) and 2.0 (late stationary phase). Equal amounts of RNA were loaded. **(C)** (*Left*) *In vitro* structure probing of *Helicobacter* 6S RNA and proposed secondary structure. 5'-end-labelled 6S RNA ($\approx$ 5 nM) was subjected to RNase T1, lead(II), RNase T2, and RNase III cleavage. Lane C: Untreated 6S RNA. Lane T1: RNase T1 ladder under denaturing conditions. The position of cleaved G residues is given left of the gel. Lane OH: Alkaline ladder. (*Right*) Proposed secondary structure of 6S RNA based on *in vitro* structure mapping. Cleavages by RNase T1, T2, or lead(II) are indicated by black, red, and blue arrows, respectively. The transcriptional start of pRNAs is indicated by a green arrow and the transcribed nucleotides are indicated in green. Location of additional pRNAs* according to Fig. 6.15 that start in the reverse direction of the central bubble are indicated in blue.

central bubble as in *E. coli* (Fig. 6.16C, green arrows). Therefore, the libraries are also suitable for detecting very short transcripts and, thus, for the first time allow detection of pRNAs *in vivo*.

6.1.7. Novel sRNAs

The current annotation of *H. pylori* genomes is largely focussed on mRNA, tRNA, and rRNA genes, and only a small number of sRNAs was identified in the biocomputational approach (Section 6.1.1).

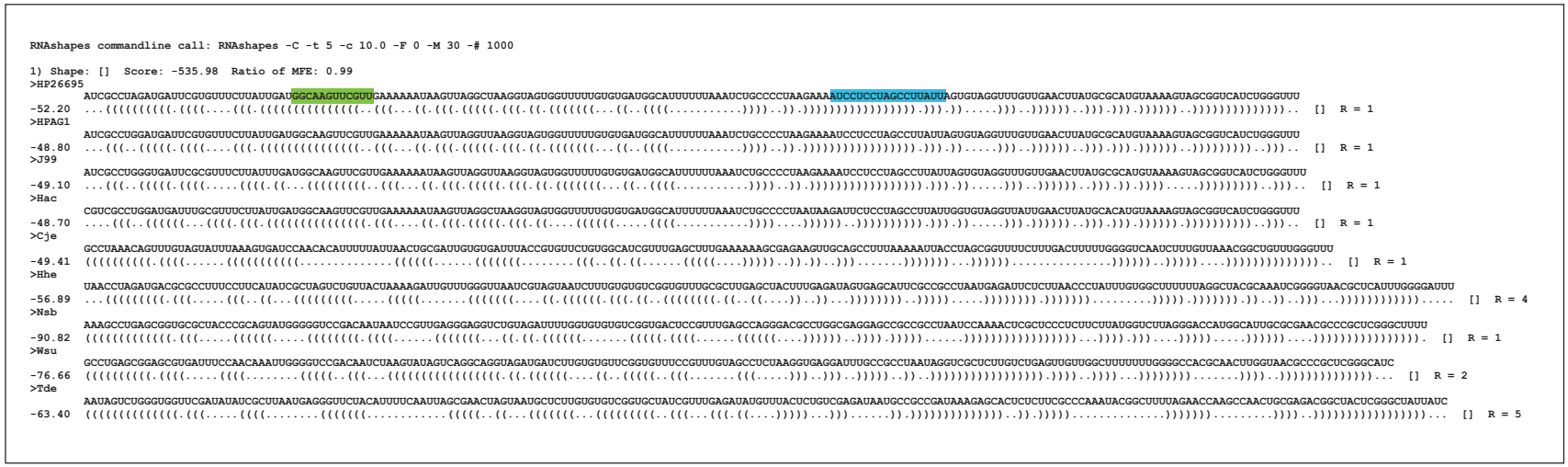


Figure 6.17: RNAshapes secondary structure of 6S homologues from diverse ϵ-proteobacteria. To identify 6S homologues in diverse ϵ-proteobacteria, BLAST searches for *purD* (HP1218) homologues were carried out, the *purD* upstream regions corresponding to the genomic location of 6S RNA in *H. pylori* 26695 extracted and manually inspected for -10 promoter elements. Based on the assumption that the 6S RNA location is conserved, 6S homologues could be identified in diverse members of the ϵ-subdivision. **HP26695:** *H. pylori* 26695 (NC_000915), **HPAG1:** *H. pylori* HPAG1 (NC_008086), **J99:** *H. pylori* J99 (NC_000921), **Hac:** *Helicobacter acinonychis* str. Sheeba (NC_008229), **Cje:** *Campylobacter jejuni* subsp. jejuni NCTC 11168 (NC_002163), **Hhe:** *Helicobacter hepaticus* ATCC 51449 (NC_004917), **Nsb:** *Nitratiruptor* sp. SB155-2 (NC_009662), **Wsu:** *Wolinella succinogenes* DSM 1740 (NC_005090), **Tde:** *Thiomicrospira denitrificans* (NC_007575). Prediction of consensus RNA shapes using RNAshapes (Reeder & Giegerich, 2005; Steffen *et al.*, 2006) showed that all putative 6S RNAs can fold into the conserved long stem-loop structure. Only the thermodynamically best shape is shown. First, the common shape type ([]), and the absolute and relative score are listed. In addition, for each input RNA, the sequence, the predicted shrep, (thermodynamically best structure which adopts this shape), its energy, and its individual rank in the shape space is given. Note that sequences are not aligned. Regions that are transcribed into pRNAs and pRNAs* are indicated in the *H. pylori* 26695 sequence in green and blue, respectively. Note that the central bubble shows additional base-pairs compared to the experimentally derived secondary structure in Fig. 6.16C, as the algorithm tries to find the thermodynamically best structure.

In contrast, the sequencing data revealed hundreds of candidate sRNA transcripts expressed from IGRs, from regions antisense to annotated ORFs, and in a few cases from the sense strand of protein coding regions. During the manual annotation of all *H. pylori* TSSs, also putative sRNA candidates were annotated. In total, ≈840 putative ncRNA TSSs were identified. These include orphan transcripts, which are putative sRNA candidates, *cis*-encoded antisense RNAs, and strong internal transcripts which are located sense to mRNA genes.

Figure 6.18 shows some examples of the different classes of novel ncRNAs that were identified in *Helicobacter pylori*. The first one, sRNA A is an independent transcript of ≈87 nt and encoded in the intergenic region between HP1044 and HP1043, whereof the latter encodes for one of the orphan response regulators (Fig. 6.18A). For some sRNA candidates it remains unclear from the 454 data whether they are independent transcripts or actually correspond to 5' UTRs of mRNA genes, and reads stop upstream of the annotation due to 454 sequencing length limitations. The next two candidates, sRNA B and D are clearly detectable as ≈142 nt and ≈213 nt long RNAs correlating with the read patterns observed in the 454 data. However, some sRNA candidates already turned out to be 5' UTRs as only a long RNA corresponding to the size of the full length mRNAs was detected on Northern blots (data not shown). Candidate sRNA D is especially interesting as it is encoded in the *cag* pathogenicity island and, thus, could have a function in virulence.

Besides separate standing sRNAs, a large fraction of *cis*-encoded antisense RNAs was detected (for examples see Fig. 6.18B). Some of the antisense RNAs are located on the opposite strand of hypothetical proteins (for example asRNAs 1a/b, 7) which could be misannotated; however, some were also found antisense to conserved proteins, *e. g.*, sRNA L which is antisense to a gene encoding for a short chain alcohol dehydrogenase. In addition, antisense transcripts to tRNAs or the 23S rRNA leader (for example asRNA 11a/b) were observed. Furthermore, antisense RNAs were found on the opposite strands of 5' or 3' UTRs or within mRNA genes. In addition, these antisense RNAs are often present in different lengths (see for example sRNA L). Besides RNAs that are located within mRNA genes on the opposite strand, several sense encoded internal sRNAs were identified (Fig. 6.18C). These sense RNAs are also often transcribed or processed into RNAs of different length and their expression seems to be growth or stress regulated, *e. g.*, ssRNA I and IIIa/b which are both strongly induced upon 30 min of acid stress. Some of the ncRNAs are duplicated in the *H. pylori* genome (for example asRNA 1a/b, asRNA 11a/b, and ssRNA IIIa/b). In total, expression of more than 30 sRNAs could be confirmed on Northern blots so far (data not shown), but the functional roles of the identified sRNAs remain to be elucidated.

6.1.8. Additional short ORFs within the *H. pylori* genome

In addition to new sRNAs, the transcripts detected in intergenic regions could also encode for small, non-annotated peptides. In the biocomputational screen, a family of spRNAs and associated asRNAs was identified. Two additional transcripts, HPnc4160 (asRNA G) and HPnc4170 (spRNA G), that were detected in the 454 data and could be validated on Northern blots (Fig. 6.19A), are located antisense to each other (Fig. 6.19B). The longer transcript, HPnc4170, could encode for a 44 aa long peptide which is conserved in several other strains (Fig. 6.19A). The smaller transcript HPnc4160 could act as a *cis*-encoded antisense RNA that represses translation of the peptide similarly to the previously identified asRNAs A to F.

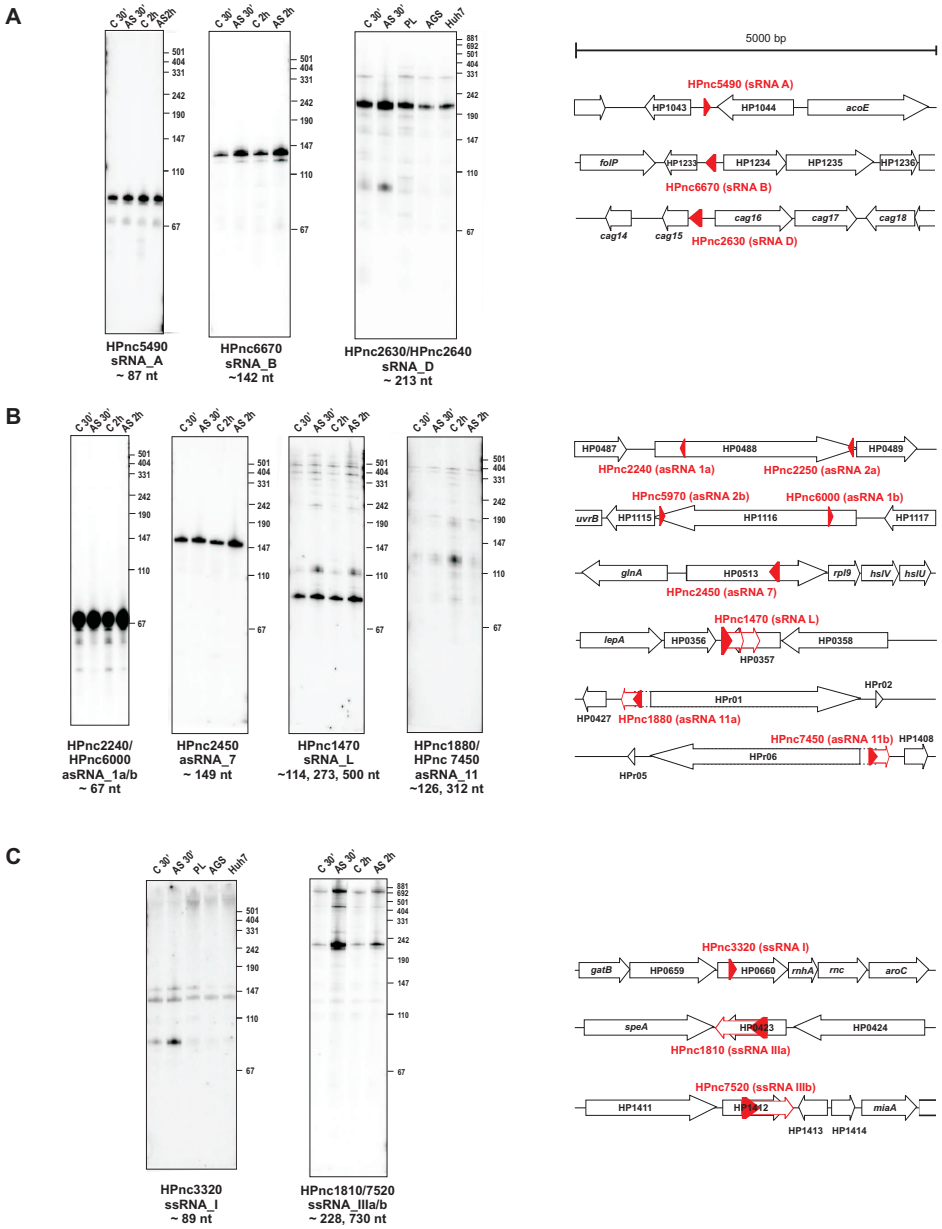


Figure 6.18: Examples of novel sRNAs in *Helicobacter*. **(A-C)** Blots showing the detection of stable transcripts of nine novel sRNAs in *H. pylori*. Total RNA was isolated from *Helicobacter pylori* at different growth or stress conditions: (C 30') culture in BHI medium to mid-log phase, reference sample for the 30' timepoint, (AS 30') 30 min acid stress pH 5.2, (C 2h) culture in BHI medium to mid-log phase, reference sample for the 2h timepoint, (AS 2h) 2 hours acid stress pH 5.2, (PL) growth in cell culture medium, (AGS) growth in the presence of AGS cells, and (Huh7) growth in the presence of Huh7 cells. 10 or 15 μg RNA of each sample were analysed on Northern blots and expression of sRNAs validated by hybridization with DNA oligonucleotide probes. **(A)** Examples for sRNAs encoded in intergenic regions, **(B)** *cis*-encoded antisense RNAs, and **(C)** sRNAs that are located internal and sense to mRNA genes. The genomic context for each sRNA (red arrows) is indicated schematically next to the blots. Note that HP0488 and HP1116, as well as HP0423 and HP1412, are gene duplications. HP0488 and HP1116 have an additional asRNA, namely, asRNA2a/b.

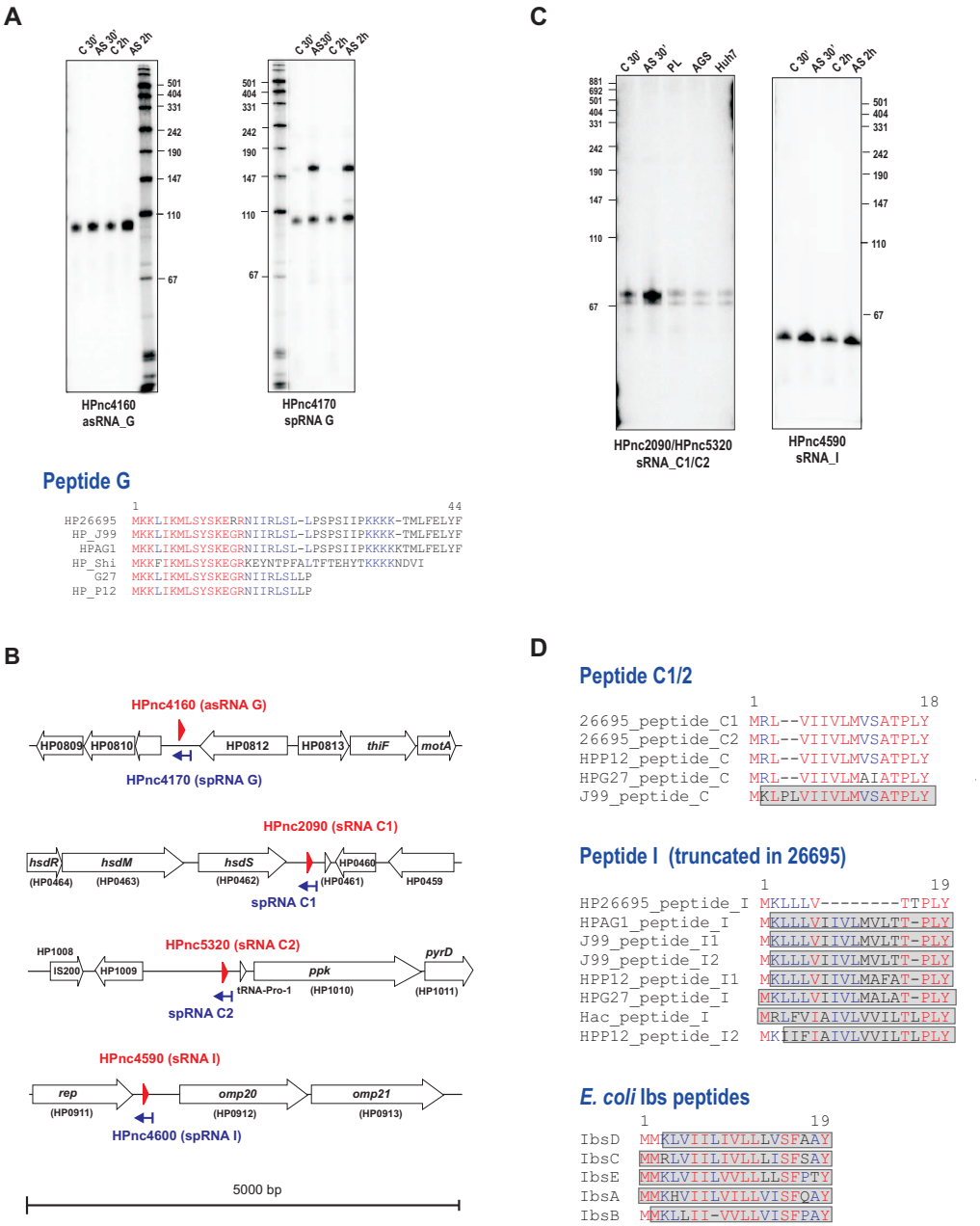


Figure 6.19: Additional spRNA/asRNA families. **(A)** Northern blots of RNA harvested from different growth conditions as in Fig. 6.18 confirmed expression of two new RNA transcripts, HPnc4160 and HPnc4170, which are encoded antisense to each other. Alignment of the putative peptide encoded by HPnc4170 and homologues that were identified in diverse *H. pylori* strains. **(B)** Genomic location of additional short-peptide encoding RNAs (blue arrows) and associated asRNAs (red triangles). **(C)** Northern blots of antisense RNAs HPnc2090/HPnc5320 and HPnc4590. **(D)** *(Top)* and *(Middle)*: Alignments of putative peptides encoded on the opposite strand of HPnc2090/HPnc5320 (peptide C1/2) and HPnc490 (peptide I) and homologues predicted in diverse *H. pylori* strains. Transmembrane domains were predicted using the `TMPred` (www.ch.embnet.orgsoftwareTMPRED_form.html) program and are shaded in grey for each protein. *(Bottom)* Alignment of the Ibs peptide family from *E. coli* adapted from Fozo *et al.* (2008a).

Furthermore, two of the newly identified sRNAs, sRNA C1/2 (HPnc2090/HPnc5320) which is 100% duplicated in the genome and sRNA I (HPnc4590) (Fig. 6.19C) contain a nearly perfect anti-Shine-Dalgarno sequence 'TCTCCT' and are highly expressed. On the opposite strands of sRNAs C1/2 and sRNA I only few reads were detected but inspection of the genomic sequences revealed that they could encode for small peptides of 16 and 11 aa (Fig. 6.19D). Conservation analysis showed that the Shine-Dalgarno sequences, start codons, and peptides are highly conserved in other *H. pylori* strains. Moreover, the spRNAs/asRNA pairs are present in different copy numbers at the same genomic locations in different strains (see Fig. 6.20A and B).

Additional cassettes that are similar in sequence to both, sRNA C1/2 and sRNA I, were identified in the other strains (Fig. 6.20C). Thus, regarding also the sequence similarity between peptides I and C1/2 they could constitute one large family. Furthermore, they have sequences similar to the short hydrophobic peptides of the Ibs family which has recently been identified in *E. coli* (Fozo *et al.*, 2008b). Similar to the peptides encoded by spRNAs A to F, also for peptides C1/2 and I transmembrane domains could be predicted (Fig. 6.20C). Overall, in addition to the spRNA A to F familiy, four additional spRNA/asRNA cassettes in the *H. pylori* 26695 genome were identified (Fig. 6.20D).

6.2. Discussion

An increasing number of regulatory RNAs is being characterized from all three domains of life. The first sRNAs have been identified in *E. coli*, but the number of bacteria where sRNAs have been identified is rapidly growing. Most of them regulate trans-encoded target mRNAs by antisense pairing mechanisms for which they require the ubiquitous bacterial RNA-binding protein, Hfq. However, regulatory sRNAs have so far evaded detection in *H. pylori*, including the exceptionally wide-spread 6S RNA, a regulator of RNA polymerase activity. Moreover, the *H. pylori* genome lacks an *hfq* gene (Sun *et al.*, 2002) and has significantly less intergenic space than sRNA-rich bacteria such as *E. coli*. These observations questioned that *H. pylori* encoded sRNAs and, therefore, the universal occurrence of bacterial riboregulators.

One the one hand, the previous paucity of transcriptional regulators of cellular adaptive responses in *H. pylori* may be explained by the rather constant gastric environment *H. pylori* faces upon host infection. On the other hand, the recent discoveries of myriads of small noncoding RNAs (sRNAs) in virtually all organisms have provided ample evidence for intricate regulation of adaptive and stress responses at the post-transcriptional level. In this thesis, only a few sRNAs could be predicted in *H. pylori* using a biocomputational approach (Section 6.1.1). This appoach was limited to prediction of orphan promoters and terminators in intergenic regions and was based on known sRNA features from *E. coli*. As *H. pylori* does not encode an Hfq homologue (Sun *et al.*, 2002), sRNAs in this bacterium might have different functions as well as sequence and secondary structure features. Thus, a direct RNomics approach including selective sequencing of RNAs enriched for primary transcripts was taken. This identified diverse sRNA genes in *H. pylori* and, in addition, allowed a global analyis of mRNA transcriptional start sites in this bacterium.

Previous RNomics screens for sRNA identification in bacteria, *e. g.*, in *E. coli*, were at that time based on the conventional Sanger sequencing method and, therefore, limited by the number of se-

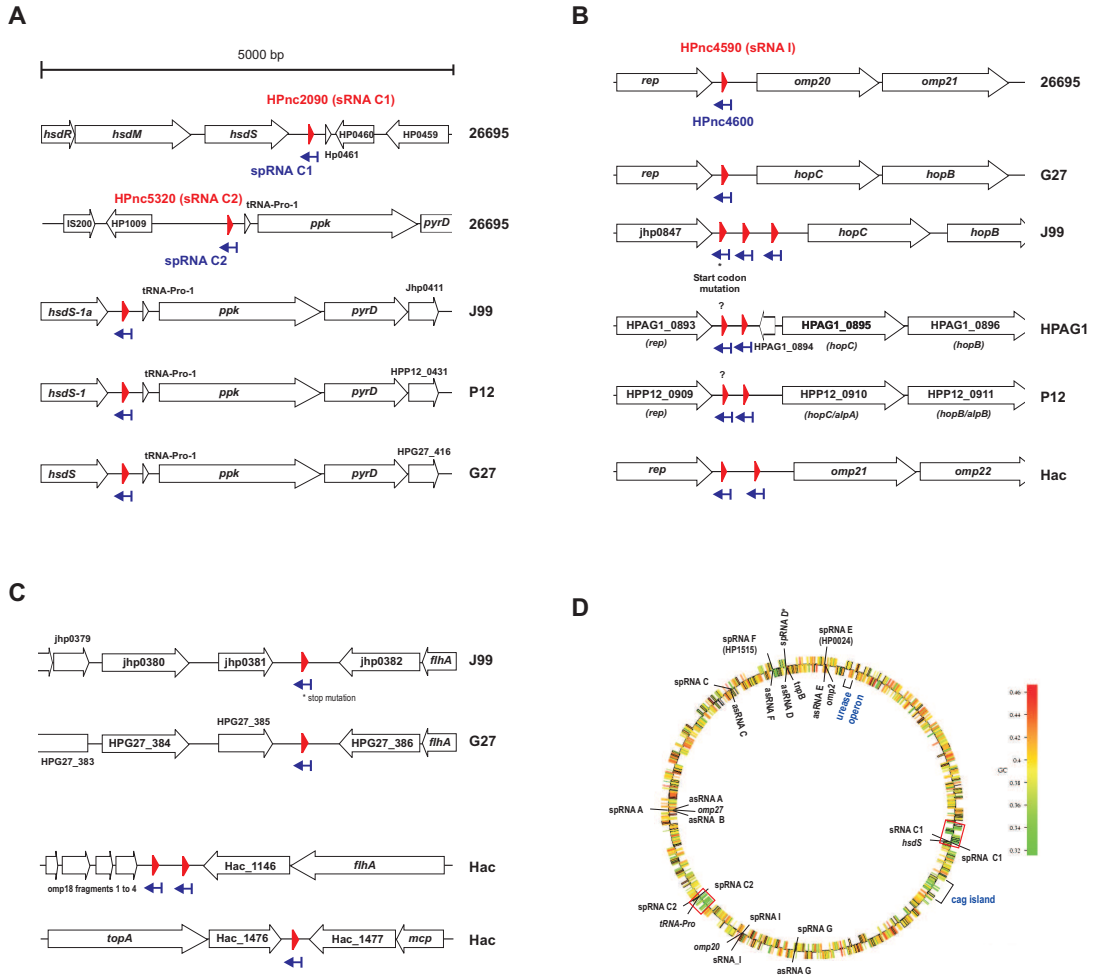


Figure 6.20: Conservation and location of diverse peptide families in *H. pylori*. **(A)** and **(B)**: Genomic locations of putative spRNA/asRNA cassettes belonging to the asRNA C1/2 and asRNA I families. Short-peptide encoding RNAs are indicated by blue arrows and associated asRNAs by red triangles, respectively. *Helicobacter* strains are given right of the locations: *H. pylori* 26695, J99, P12, HPAG1, and G27, and *H. acinonychis* (Hac). **(C)** Additional cassettes that are similar to the asRNA C1/2 and asRNA I families but encoded at different genomic locations were identified in *H. pylori* J99, *H. pylori* G27, and *H. acinonychis*. **(D)** Locations of spRNA/asRNA cassettes on the *H. pylori* 26695 chromosome. The *cag* island and urease operon are indicated in blue, the two plasticity zones (Alm *et al.*, 1999) by red boxes. The circular representation of the 26695 chromosome (taken from CampyDB: (http://xbase.bham.ac.uk/campydb/) indicates the G/C content for each gene based on the color code shown on the right.

quenced cDNA clones (see Section 2.3.5). In contrast, next-generation sequencing methods have enabled the recent 'RNA sequencing' (RNA-Seq) technology, which generates tens of millions of short sequences of random cDNA libraries in a single assay but has not been applied to bacteria so far (reviewed in Wang *et al.*, 2009). Two initial studies applied deep sequencing to identify novel genes in bacteria (Frias-Lopez *et al.*, 2008; Mao *et al.*, 2008). In this study, RNA-Seq using 454 pyrosequencing was applied to the transcriptome of *H. pylori*. During preparation of this thesis, two studies which also applied deep sequencing to transcriptome analysis in bacteria were pub-

lished (Liu *et al.*, 2009; Yoder-Himes *et al.*, 2009). Liu *et al.* (2009) used a tRNA and 5S rRNA depletion treatment which is based on RNase H digestion of RNA bound to complementary DNA oligonucleotides and analysed size-fractionated RNA of *Vibrio cholerae* by 454 sequencing. They identified 500 new putative intergenic sRNAs, 127 putative asRNAs and also transcripts from the sense strand within ORFs. The second study used Illumina RNA-Seq to analyse the transcriptome of two *Burkholderia cenocepacia* strains under two relevant environmental conditions. Here, rRNAs were depleted using the MicrobeExpress kit (Ambion). This kit is also based on capture oligonucleotides that bind to the bacterial 16S and 23S rRNAs; however, the rRNA hybrids are removed from the solution using derivatized magnetic microbeads instead of an RNase H digestion. Besides the identification of 13 new putative sRNAs in *Burkholderia*, several genes that are strain-specifically induced in one of the analysed conditions were identified giving a first glimpse at the application of transcriptional profiling using deep sequencing in bacteria. However, both approaches did not provide a global annotation of mRNA TSSs. The approach based on enrichment of primary transripts described in this thesis requires no size-fractionation step and no organism specific antisense oligos for tRNA/rRNA depletion. In parallel, also mRNA TSSs can be enriched, and this strategy should in principle also be applicable to eukaryotic mRNAs where the cap structure also prevents degradation by TEX.

In comparison to the bioinformatics-based approach, the RNA-Seq approach led to the identification of hundreds of sRNA candidates in *H. pylori*, whereof more than 30 could be verified on Northern blots so far. Diverse sRNAs located in intergenic regions as well as sense or antisense to annotated genes were detected. For some sRNA candidates it is unclear from the 454 sequencing data whether they are independent transcripts or actually correspond to 5' UTRs of mRNA genes as reads stop upstream of the annotation due to 454 sequencing lengths limitations. This requires further experimental investigation of actual transcript lengths on Northern blots. Recently, two studies tried to identify sRNAs in *H. pylori* by a bioinformatics-based approach (Xiao *et al.*, 2009a) and antisense transcripts by an experimental approach based on an RNase I protection assay (Xiao *et al.*, 2009b). However, they identified only two sRNAs and two antisense RNAs, but none of them was recovered in the 454 sequencing data.

Among the novel sRNAs, a homologue of the highly abundant 6S RNA was identified which previous bioinformatics-based approaches failed to detect (see Section 2.1.3.1 and Barrick *et al.*, 2005; Willkomm *et al.*, 2005). It is encoded upstream of *purD*, and the same region was previously suggested to be a riboswitch based on comparative structure analysis (Weinberg *et al.*, 2007). In contrast, the experimentally verified conserved 6S structure as well as detection of the pRNAs in the 454 data that are specifically transcribed from 6S RNA argue that this region encodes a 6S homologue. In *E. coli*, it was shown that 6S RNA facilitates transcription of σ^S-dependent promoters by titrating σ^{70}-bound RNA polymerase (Wassarman, 2007). So far, no homologue of σ^S is known in *H. pylori*. However, it was recently reported that promoters with a weak -35 element are senstitive to 6S RNA regulation and that an extended -10 element similarly determines inhibition by 6S RNA except when a consensus -35 element is present (Cavanagh *et al.*, 2008). As there is no common -35 element known in *H. pylori*, an extended set of genes could be sensitive to 6S RNA regulation also in this bacterium.

Besides 6S RNA and sRNAs in intergenic regions, a large number of antisense transcripts were identified in *H. pylori* which are located antisense to 5' or 3' UTRs or within open reading frames.

These *cis*-encoded antisense RNAs probably base-pair with the RNAs encoded on the opposite strand and thereby inhibit translation of their targets or lead to degradation by the double-stranded specific RNase III. Especially, *cis*-encoded antisense RNAs to 5' UTRs could act by inhibition of translation as described for SymR RNA in *E. coli*, which is transcribed opposite to the 5' end of the SOS-induced toxin SymE (Kawano *et al.*, 2007). In cyanobacteria, an internal antisense RNA was described to control expression of a photosystem associated protein in response to environmental changes (Dühring *et al.*, 2006). Therefore, the diverse antisense transcripts could also allow rapid regulation of gene expression in *Helicobacter* upon different stresses. Recently, transcription of the *ubiGmccBA* operon of *Clostridium acetobutylicum*, which is involved in methionine to cysteine conversion, was shown to be controlled by a cysteine-specific T-box riboswitch in the 5' UTR but mainly by several antisense RNAs which are themselves controlled by an S-box riboswitch (André *et al.*, 2008). The abundance of sense and antisense transcripts was inversely correlated with the sulfur source availability, and the antisense RNAs were shown to modulate the level of *ubiG* transcript and of MccB activity. Alternatively, the antisense RNAs could lead to 3' or 5' end processing of mRNAs and alter transcript stabilities as shown for GadY sRNA in *E. coli* (Opdyke *et al.*, 2004). Specifically, the *gadY* gene was shown to overlap the 3' end of the *gadX* gene, and this overlapping region was found to be necessary for the GadY-dependent increase of *gadX* mRNA stability. In *Vibrio anguillarum*, a plasmid-encoded antisense RNA leads to transcription termination after the *fatA* gene of the *fatDCBAangRT* operon and, hence, reduces expression of the downstream *angRT* genes (Stork *et al.*, 2007).

As most of the computational screens are focussed on the identification of independent sRNA genes in intergenic regions they will miss sRNAs which are located antisense or internal to annotated genes (see Section 2.3.3). In *E. coli*, a global low level antisense transcription has been previously reported based on a whole-genome microarray (Selinger *et al.*, 2000). Moreover, a cloning-based screen of small RNAs $<$65 nt identified several *cis*-encoded antisense RNAs as well as 5'- and 3'-UTR-derived small RNAs in *E. coli* (Kawano *et al.*, 2005a). Meanwhile, this approach was limited to the conventional Sanger sequencing similar to the first RNomics screens. Thus, a higher sequencing depth could reveal additional *cis*-encoded sRNA genes. In a second approach, the same group detected promoter activities within open reading frame sequences in *E. coli* by cloning random fragments upstream of a promoterless *lacZ* gene on a plasmid and measuring β-galactosidase activities (Kawano *et al.*, 2005b). However, they failed to detect defined transcripts for the identified promoters on Northern blots and suggested that these promoters do not give rise to stable transcripts. Although this indicates that it is difficult to identify and detect such transcripts in global approaches, the deep sequencing strategy presented in this Chapter successfully identified many antisense RNAs and sRNAs sense to ORFs.

The first bacterial antisense RNAs were actually identified on plasmids (see Section 2.1.5). Some of these *cis*-encoded plasmid sRNAs repress the synthesis of toxic proteins and act as plasmid-addiction molecules or post-segregational killing systems (Gerdes *et al.*, 1997). In this case, a very unstable RNA inhibits translation of a very stable mRNA that encodes the toxin. As soon as the plasmid is lost, the unstable asRNA gets quickly degraded which allows translation of the more stable mRNA, and the toxin produced kills the bacteria. The best known toxin-antitoxin pair is the *hok/sok* system of plasmid R1 of *E. coli* (Gerdes *et al.*, 1990). The biocomputational prediction and the deep sequencing approach identified several potential asRNA/short-peptide encoding RNA

cassettes in the *H. pylori* genome. For example, asRNAs A to F and the associated spRNAs constitute a family which is present in different copy numbers in diverse *Helicobacter* strains. Some of these copies show mutations, indicating that they could have a redundant function and not all are required at the same time. The hydrophobic and positively charged peptides that are encoded by the spRNAs could interact with membranes and thereby act as chromosomally encoded toxins. Recently, several small toxic proteins that are present in multiple copies in the chromosome were identified in *E. coli* and were found to be repressed by *cis*-encoded antisense RNAs (Fozo *et al.*, 2008b). In addition, there is growing evidence that short ORFs have been largely overlooked in *E. coli* and *S. typhimurium* during genome annotations (Alix & Blanc-Potard, 2009; Fozo *et al.*, 2008a; Hemm *et al.*, 2008). As the short peptides in *Helicobacter* could probably act as toxins, it can be difficult to detect them *in vivo*. At least in an *in vitro* translation system it could be shown that they are effectively translated and that a 1:1 ratio of the corresponding asRNA is sufficient to repress translation of the peptide (F. Darfeuille, unpublished data). In total, at least three different families of putative short peptide/asRNA cassettes in the 1.67 Mb *Helicobacter* genome were identified, which fits well with the number of at least six families described in the 4.6 Mb *E. coli* (Alix & Blanc-Potard, 2009; Fozo *et al.*, 2008a).

Interestingly, some of the asRNA/peptide cassettes are located close to the highly variable plasticity zones of *Helicobacter* (Alm *et al.*, 1999; Gressmann *et al.*, 2005) and could have a potential function in genome integrity. These plasticity regions show a lower G/C content than the rest of the genome and contain several insertion sequences. Based on the following observations, it was suggested that they were aquired by horizontal transfer from plasmids (Alm *et al.*, 1999). First, *H. pylori* and *Campylobacter* spp. plasmids have a G/C content in this lower range and, second, two copies of the insertion-sequence element as well as neighbouring 26695-strain specific chromosomal DNA from the plasticity zones are present in *H. pylori* plasmid pHPM186 (Lee *et al.*, 1997).

The newly identified peptides could have a role in altruistic autolysis of *Helicobacter* which was suggested as a mechanism for the release of several cytoplasmic proteins (Dunn *et al.*, 1997; Marcus & Scott, 2001; Phadnis *et al.*, 1996; Vanet & Labigne, 1998). For example, the presence of urease both in the cytoplasm and bound to the outside surface of the bacteria was observed *in vitro* (Phadnis *et al.*, 1996) and *in vivo* in human gastric biopsies (Dunn *et al.*, 1997). Fujita *et al.* (2005) observed that autolysis occurred after late-log phase of culture and identified a heat-stable and hydrophobic peptidergic fraction of $< 3{,}500$ Da as the autolysis-inducing factor (AIF). The partially purified AIF had a lytic activity which is specific for *H. pylori* and *Campylobacter jejuni*, but not, *e. g.*, for *E. coli* and *Salmonella*. Maybe some of the newly identified hydrophobic peptides are part of this AIF.

The ubiquitous RNA-binding chaperone Hfq is required for the action of most of the enterobacterial sRNAs that regulate trans-encoded target mRNAs by antisense pairing mechanisms. So far, no Hfq homologue has been identified in *H. pylori*. Therefore, the Hfq-coIP strategy from the previous Chapter that was applied to detect novel sRNAs in *Salmonella* (Sittka *et al.*, 2008) could not be applied for sRNA identification in *Helicobacter*. Nevertheless, the question remains whether the newly identified sRNAs that were identified by the deep sequencing approach require RNA binding proteins or chaperones for their function or stability. To screen for auxiliary RNA-binding proteins in *H. pylori*, aptamer tags could be added to several of the new *H. pylori* sRNAs and sRNA-binding proteins isolated by affinity chromatography similar to previously described approaches in *E. coli* or *Salmonella* (Windbichler *et al.*, 2008, and Said *et al.*, 2009, submitted). Another approach would be

to co-immunoprecipitate RNAs with epitope-tagged putative RNA-binding proteins from *H. pylori* as decribed for the Hfq-coIP in *Salmonella* in the previous Chapter. One interesting candidate is an 82 aa-long RNA binding protein (HP0827) which is also present in various species of cyanobacteria and *Treponema pallidum* but absent from almost every other completely sequenced prokaryotic genome (Maruyama *et al.*, 1999). This protein contains a single RNA recognition motif (RRM) which is also found in several eukaryotic RNA-binding proteins involved in splicing or control of mRNA translation. Interestingly, also in the cyanobacterium *Prochlorococcus* strain MED4 several sRNA and antisense RNAs have been identified despite a lack of an Hfq homologue (Axmann *et al.*, 2005; Steglich *et al.*, 2008). It was suggested that maybe novel mechanisms for RNA-RNA interactions may exist in this group. It would be interesting to see if any of the new sRNAs and antisense RNAs that were identified in *H. pylori* or cyanobacteria are specifically bound by this protein.

In addition to the identification of diverse sRNAs, the enrichment of primary transcripts allowed a global annotation of $\approx$800 transcriptional start sites of mRNAs in *H. pylori* (Section 6.1.5). As *Helicobacter* encodes only 1,576 ORFs and genes are often transcribed as polycistronic mRNAs in bacteria, probably most of the mRNA TSSs were covered. In *Helicobacter*, not much is known about promoter signals as well as transcription factor binding sites, and biocomputational promoter predictions identified only a limited number of promoter motifs (Vanet *et al.*, 2000). Based on the global transcription map defined in this study, it is now possible to extract regions upstream of the TSSs and to try to identify common promoter motifs or special promoter patterns for functional classes of genes, for example using `MEME` (Bailey *et al.*, 2006). Preliminary analysis identified a highly conserved -10 box (TATAAT) but instead of a -35 box a very strong periodic variation in the AT-content and semi-conserved T-stretches, with a period of 10-11 nucleotides (data not shown) similar to what was reported for *Campylobacter* (Petersen *et al.*, 2003).

Furthermore, the mRNA TSS map allowed calculating the UTR lengths distribution in *H. pylori* (Fig. 6.14). Most of the genes were preceded by UTRs of 20-50 bp, whereas almost no shorter UTRs, except for leaderless mRNAs, could be identified. This is in accordance with the region known to be covered by the 30S subunit during translation initiation ranging from $\approx$-20 bp upstream of the start codon to $\approx$+19 bp downstream of the start codon as derived from biochemical footprinting experiments (Beyer *et al.*, 1994; Hüttenhofer & Noller, 1994). Morerover, X-ray analyses showed that a region of $\approx$30 nt involving the SD sequence is wrapped around the 30S subunit (Yusupova *et al.*, 2006, 2001). In *E. coli*, mRNAs with a weak SD sequence usually carry a pyrimidine-rich region 5' to the SD that acts as a recognition motif for the ribosomal protein S1 and anchors the mRNA on the ribosome (Boni *et al.*, 1991; Komarova *et al.*, 2005). Thus, in very short UTRs ($<$20 bp) there would not be enough space for a ribosome binding site. In addition, several quite long UTRs up to 500 bp (which was the cut-off for the distance of a primary or secondary TSS) were identified. These could potentially harbour regulatory elements and, for example, secondary-structure clustering (Will *et al.*, 2007) with known riboswitches from, *e. g.*, *Bacillus subtilis* and *E. coli*, or comparison of RNA shapes with `Rfam`[4] families (Janssen *et al.*, 2008) could be used to identify known riboswitch types.

[4] `www.sanger.ac.uk/Software/Rfam/`

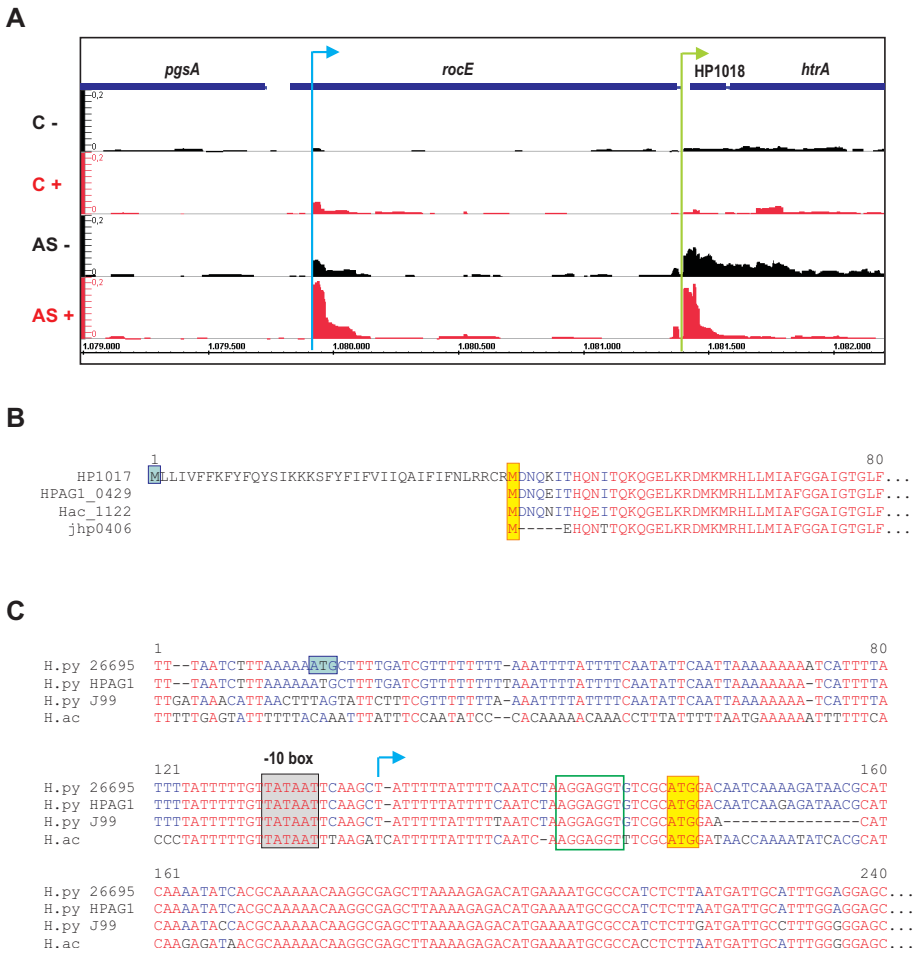


Figure 6.21: Reannotation of *rocE* based on deep sequencing. **(A)** Screen shot of the IGB showing the read distribution of the C-/+ and AS-/+ libraries on the plus strand in the genomic region harbouring *pgsA*, *rocE*, HP1018, and *htrA*. Transcription of *rocE* starts within its annotated ORF (blue arrow), and the downstream gene HP1018 is transcribed from an additional TSSs (green arrow). **(B)** Protein alignment of the N-terminal sequences of RocE homologues from diverse *H. pylori* strains shows that the N-terminus is shorter in the other strains. The probably correct start position of the protein is marked by a yellow box. **(C)** Nucleotide alignment of promoter regions and 5' parts of *rocE* genes from dfferent *H. pylori* strains. The annotated start codon in *H. pylori* 26695 is marked by a blue box, the suggested start codon for reannotation by a yellow box, respectively. The -10 promoter element and the transcriptional start site based on the 454 data are shown in grey and by a blue arrow, respectively. A potential Shine-Dalgarno sequence upstream of the suggested correct start codon is marked in green.

During the analysis of UTR lengths, 25 mRNAs were found to be leaderless in *H. pylori*. In contrast to translation initiation by 30S binding on mRNAs which contain a 5' UTR, translation of leaderless mRNA was shown to be initiated by the assembled 70S ribosome and thereby bypassing the dissociation process (Moll *et al.*, 2004; Udagawa *et al.*, 2004). Ribosome binding assays revealed that a leaderless mRNA's 5'-AUG is required for stable binding to the ribosome and that addition of a 5'-terminal AUG triplet to a random RNA fragment can make it both competent and competitive for ribosome binding (Brock *et al.*, 2008). Indeed, the transcriptional start sites of the 25 leader-

less mRNAs in *H. pylori* exactly mapped to the 'A' of the AUG start codon. Leaderless mRNAs appear to be rather infrequent in Gram-negative bacteria. These include, *e. g.*, the λcI mRNA (Walz *et al.*, 1976) and the Tn*1721 tetR* mRNA (Baumeister *et al.*, 1991) derived from accessory genetic elements in *E. coli*. Their number in different Gram-positive genera such as *Streptococci*, *Lactococci*, *Streptomyces* and *Corynebacterium* by far exceed that identified in Gram-negative species, and leaderless mRNAs are quite common in archea (reviewed in Moll *et al.*, 2002). In *Helicobacter*, several hypothetical genes were found to be transcribed as leaderless mRNAs but also those of some proteins with important cellular functions, such as DnaA, RecR, and HemH (see Table 10.13). The functional relevance of why these genes are leaderless, however, remains to be identified. In *E. coli* is has been shown recently that prolonged exposure of ribosomes to the translation-initiation inhibiting antibiotic kasugamycin triggers the spontaneous loss of small subunit proteins and produces a reduced ribosomal particle that exclusively translates leaderless mRNAs (Kaberdina *et al.*, 2009).

The analysis of transcriptomes by deep sequencing is not only appropriate for identification of novel sRNAs and mapping of transcriptional start sites but can also help to reannotate genomes based on actual transcription of genes. During the manual TSS annotation, it was observed that transcription for several mRNAs started downstream of the annotated start codon, and no convincing TSS upstream of the ORF was observed (see for example Figures 6.21 and 6.22). This raised the question, whether the start codons of these proteins are misannotated and that translation actually starts downstream of the transcriptional start site. Conservation analysis of protein and nucleotide sequences of, *e. g.*, HP0112 and *rocF* indicated that the N-termini of these proteins are actually shorter in diverse *H. pylori* strains and that they have to be reannotated in *H. pylori* 26695 (Figs. 6.21 and 6.22, B and C). Furthermore, no promising SD-sequence is located upstream of the old start codon of *rocF*, whereas a perfect and highly conserved SD sequence is located upstream of the proposed correct start codon (Fig. 6.21C). In case of HP0112, the protein starts with an alternative start codon 'TTG' which is not conserved in the other strains. Based on the 454 data, the ORF starts on the 'A' of the proposed new start codon leading to a leaderless mRNA. Interestingly, immediately downstram of the reannotated start codon a C/A rich element is located. These elements have been shown to act as translational enhancers of leadered and unleadered mRNAs in *E. coli* (Martin-Farmer & Janssen, 1999). In total, about 19 ORFs were identified that probably have to be reannotated based on the 454 data and conservation analysis (data not shown). Most of them are also included in a previously published list of genes for a revised annotation based on comparison between the two strains 26695 and J99 (Boneca *et al.*, 2003). Thus, the global mapping of the actual transcriptional start sites of genes could help to identify the correct start codons of ORFs and to reannotate bacterial genomes based on actual gene expression data. The application of deep sequencing for genome annotations has been successfully used also in eukaryotes (Weber *et al.*, 2007; Yassour *et al.*, 2009). Thus, in future, probably not only DNA but also RNA will be sequenced for improved genome assembly and subsequent annotation of bacterial genomes.

In bacteria, the 5' end of newly synthesized RNAs bears a triphosphate derived from the first transcribed nucleotide. Thus, the approach for selective sequencing primary transcripts to map mRNA transcriptional start sites or to identify novel sRNA is not limited to *H. pylori* and could, in principle, be applied to diverse bacteria. Moreover, it should generally be applicable also to eukaryotic mRNAs where the distal phosphate is replaced by an inverted, methylated Guanosine-

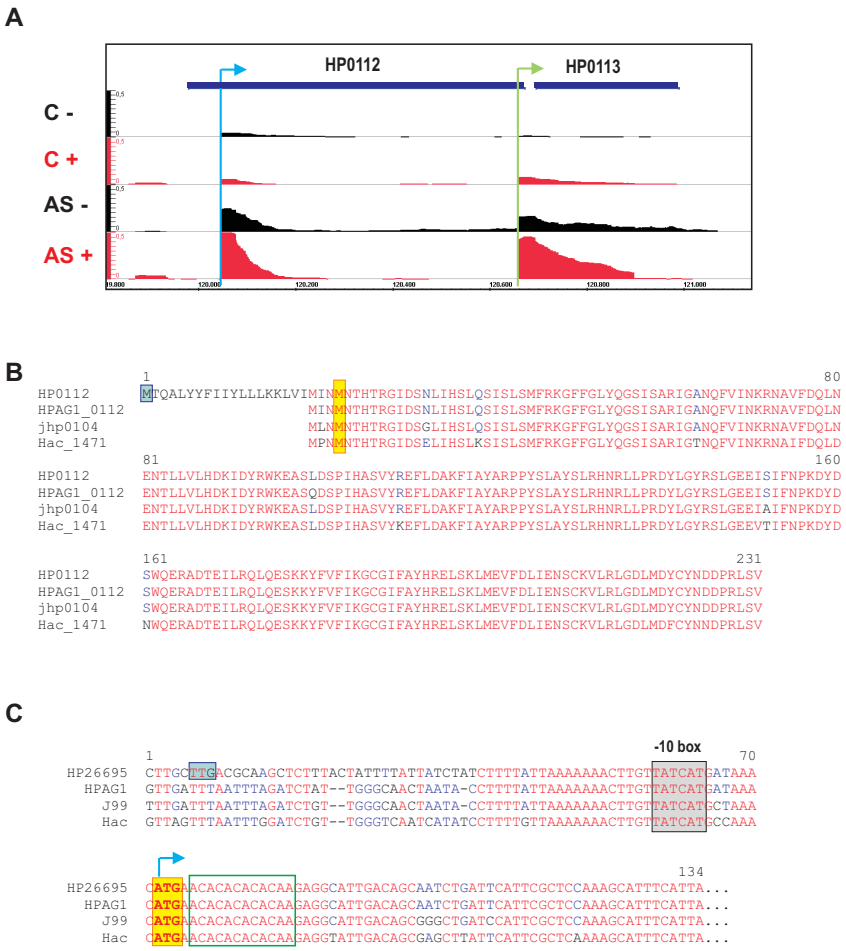


Figure 6.22: Reannotation of HP0112 based on deep sequencing. (A) Screen shot of the IGB showing the read distribution on the plus strand of the C-/+ and AS-/+ libraries on two hypothetical proteins, HP0112 and HP0113. Transcription of HP0112 starts within its annotated ORF (blue arrow), and the downstream gene HP0113 is transcribed from an additional TSSs (green arrow). **(B)** Protein alignment of HP0112 homologues from diverse *H. pylori* strains shows that the N-terminus is shorter in the other strains. Furthermore, the protein starts with an alternative start codon 'TTG' in *H. pylori* 26695 which is not conserved in the other bacteria as shown in **(C)**. The probably correct start position of the protein is marked by a yellow box. **(C)** Nucleotide alignment of HP0112 homologues from different *H. pylori* strains. The annotated start codon in *H. pylori* 26695 is marked by a blue box, the suggested start codon for reannotation is marked in yellow. The -10 promoter element and the transcriptional start site based on the 454 data are shown in grey and by a blue arrow, respectively. A C/A rich element downstream of the suggested correct start codon is marked in green.

monophosphate to form the m^7GpppX cap and prevents degradation by TEX. Therefore, this approach could be helpful in global TSS annotation of diverse bacteria as well as eukaryotes and, thus, help to refine current annotations.

CHAPTER 7

CONCLUSIONS

In this thesis, two aspects of sRNA-mediated regulation in bacteria have been investigated: multiple target regulation and approaches for the identification of novel sRNAs in bacteria. In the first part, the functional characterization of a multiple-target regulator, GcvB RNA, of *Salmonella*, was presented. Proteomics and bioinformatics-based approaches led to the identifcation of seven ABC transporter mRNAs as GcvB targets. Alignment of GcvB homologues of distantly related bacteria revealed a conserved G/U-rich element, R1, that is strictly required for GcvB target recognition. Analysis of target gene fusion regulation *in vivo*, and *in vitro* structure probing and translation assays showed that GcvB represses its target mRNAs by binding to extended C/A-rich regions inside and upstream of the ribosome binding site. The presence of the C/A rich element - which can act as a translational enhancer - in all ABC transporter mRNAs indicated that this small RNA has hijacked a conserved element to regulate a class of functionally related genes. Pulse-expression of GcvB wild-type and mutant RNAs followed by global analysis of mRNA changes on microarrays led to the identification of additional R1-dependent GcvB targets and one R-1 independent target, *cycA*. The definition of a consensus motif for the C/A-rich target sites could be successfully used to refine bioinformatics-based target-predictions using `RNAhybrid` (Rehmsmeier *et al.*, 2004). These predictions revealed additional GcvB targets, which might have escaped the microarray approaches due to low expression or only slight changes on mRNA levels. Overall, GcvB RNA was found to directly regulate multiple genes involved in amino acid transport or metabolism. Furthermore, a highly conserved region turned out to be critical for regulation.

This work and several other recent sudies have revealed multiple-targeting as a widespread mode of action for diverse bacterial sRNAs (Section 2.5). The presence of conserved, single-stranded elements in sRNAs mediating multiple target recognition of frequently functionally related genes now seems to be a common denominator. Furthermore, mRNA regions distant from Shine-Dalgarno (SD) sequences can also be targeted by sRNAs as revealed in this and other studies. Therefore, future target-prediction programs could probably be improved by including the following criteria: (1) identification of conserved sRNA parts to limit the search space within the input sequence, (2) incorporation of mRNA regions outside the ribosome binding site, and (3) scoring for functional relation between targets. Furthermore, it could be helpful to include criteria such as target-site accessibility, Hfq-binding and/or prediction of Hfq-binding sites as previously suggested (Busch *et al.*, 2008; Tafer & Hofacker, 2008; Zhang *et al.*, 2006).

In the other part of this thesis, deep sequencing was applied to analyse RNA ligands bound to *Salmonella* Hfq, one of the key players in sRNA-mediated regulation. A bioinformatics-based work flow that was developed for the analysis and visualization of the deep sequencing data was presented. The identification of novel sRNAs in *Salmonella* as well as more than 700 mRNAs

bound to Hfq showed that this approach is capable of identifying sRNAs specific to certain bacteria and of limiting the mRNA input-set for target prediction programs. In furture, incorporation of a cross-linking step and digestion of RNA regions that are not covered by Hfq could narrow down the regions that are actually bound by Hfq. This could help to define Hfq consensus binding sites which could in turn be incorporated into target-prediction algorithms. Although this approach is limited to bacteria where Hfq or any other RNA binding proteins are known, it could be helpful to unravel the post-transcriptional regulons of a wide range of bacteria.

In *Helicobacter pylori* neither Hfq nor any sRNAs were previously known and the bioinformatics-based approach which was presented in the last part of this thesis identified only a few sRNAs in this organism. In contrast, a high-throughput 'RNA-seq' approach of total *H. pylori* RNA revealed many additional sRNA candidates including antisense transcripts to ORFs in *Helicobacter*. Among >30 sRNAs which have been verified on Northern blots, a homologue of the ubiquitous 6S RNA and its associated pRNAs were identified. Moreover, differential analysis of primary and processed RNA species facilitated the identification of ≈ 800 transcription start sites of mRNAs across the *H. pylori* genome. Besides a refinement of the functional annotation of the *H. pylori* and related genomes, this approach should facilitate the global transcriptome analysis of mixed pathogen-host populations as it avoids the problem of cross-hybridization observed in microarray experiments. Furthermore, initial studies suggested the use of high-throughput sequencing for transcription-profiling (Mortazavi *et al.*, 2008; Yoder-Himes *et al.*, 2009). Ongoing improvements of next-generation sequencing methods including generation of more sequences per run as well as longer average read lengths will allow to cover longer transcripts and sequencing of diverse growth conditions. In addition, the development of single-molecule sequencing technologies (Pacific Biosiences and Heliscope of Helicos) will avoid a potential bias introduced during cDNA construction in the PCR amplification or adapter ligation step and allow a more precise transcription profiling.

Current high-throughput sequencing approaches indicate the demand for appropriate databases to store sequence data as well as analysis and visualization tools. Especially the manual annotation of TSS is feasible only for small genome sizes and, hence, algorithms for an automatic recognition of enrichment patterns could be very helpful. Deep sequencing in combination with proper analysis tools will then allow a fast and effective way to improve genome annotation based on actual transcription, global promoter maps, as well as the identification of novel transcripts including regulatory RNAs.

CHAPTER 8

MATERIAL AND METHODS

8.1. Material

Chemicals used in this study were purchased from Merck (Darmstadt), Roth (Karlsruhe) and Sigma (München). The following tables list all labware (Table 8.1), instruments (Table 8.2), enzymes, proteins, and size markers (Table 8.3), and commercially available systems (Table 8.4) that were used throughout this study.

Table 8.1: Labware and Manufacturer.

Labware	**Manufacturer**
CampyGen™	Oxoid, Basingstoke, Hampshire, England
Corex II centrifuge tubes	Krackeler Scientific, Albany NY, USA
Gene Pulser/MicroPulser Cuvettes, 1 mm	BioRad, München
Horizontal Electrophoresis Systems PerfectBlue Mini S, M, L	Peqlab, Erlangen
Hybond-XL Membrane for Nucleic Acid Transfer	GE Healthcare, München
Glass Beads, 0.1 mm	Roth, Karlsruhe
Imaging Plates BAS-IP MS 2325, 2340	Fujifilm, Düsseldorf
Imaging Plates Cassettes BAS 2325, 2340	Fujifilm, Düsseldorf
Inoculation Loops 10 μl	VWR, Darmstadt
L-shape Bacteriology Loops	VWR, Darmstadt
MicroSpin G-25, G-50 Columns	GE Healthcare, München
Nucleic Acid Sequencing Unit #SG-400-20	CBS Scientific, Del Mar CA, USA
Phase Lock Gel™ (PLG) Tubes	VWR, Darmstadt
Pipetman P10, P20, P200, P1000	Gilson, Bad Camberg
Pipetboy acu	Integra Biosciences, Fernwald
PolyScreen PVDF Transfer Membrane	PerkinElmer, Waltham MA, USA
Protein A Sepharose Beads	Sigma-Aldrich, Taufkirchen
Reagent and Centrifuge Tubes 15, 50 ml	Sarstedt, Nümbrecht
Safe-Lock Tubes 1.5 ml, 2.0 ml	Eppendorf, Wesseling-Berzdorf
Semi-dry Electroblotter SEDEC M	Peqlab, Erlangen
Semi-micro Cuvettes	Sarstedt, Nümbrecht
Serological Pipets 5 ml, 10 ml, 25 ml	Corning, Wiesbaden

continued on next page

Labware	Manufacturer
Tank Electroblotter PerfectBlue Web S, M	Peqlab, Erlangen
Thermo-Tubes 0.2 ml	Abigene, Hamburg
Ventilation Cap Tubes 13 ml	Sarstedt, Nümbrecht
Vertical Electrophoresis Systems PerfectBlue Twin S, L	Peqlab, Erlangen
Sterile filters (0.2 μm pore size)	Whatman, Dassel

Table 8.2: Instruments.

Instrument	Manufacturer
Analytical Balances TE64, TE601	Sartorius, Göttingen
BD FACSCanto^{TM} Flow Cytometer	BD Biosciences, Franklin Lakes NJ, USA
Bio-Link BLX 254 UV-Crosslinker	Peqlab, Erlangen
Centrifuge 5415R, 5810R	Eppendorf, Wesseling-Berzdorf
Centrifuge RC5C Plus (Rotor: SS-34)	Thermo Scientific, Langenselbold
DNA Engine Thermal Cycler	BioRad, München
E. coli Pulser	Bio-Rad, München
Electrophoresis Power Supplies EV 232, EV202, E802	Consort, Turnhout, Belgium
Eraser for Imaging Plates	Raytest, Straubenhardt
Gel Documentation System Gel Doc 2000	Biorad, München
Gel Dryer Model 583	Biorad, München
GenePix 4000A scanner	Molecular Devices, Sunnyvale CA, USA
Hybridization Oven Compact-Line OV4	Biometra, Göttingen
Imaging System LAS-3000	Fujifilm, Düsseldorf
Incubator Innovens Category 1	Jouan, Unterhaching
Incubator Shaker Innova 44	New Brunswick Scientific, Nürtingen
MultiTempIII Thermostatic Circulator	Amersham Biosciences, Freiburg
Refrigerated Incubator Shaker C24KC	New Brunswick Scientific, Nürtingen
Spectrophotometer NanoDrop ND-1000	Peqlab, Erlangen
Thermomixer comfort	Eppendorf, Wesseling-Berzdorf
Phosphoimager FLA-3000	Fujifilm, Düsseldorf
Ultrospec 10 photometer	Amersham Biosciences, Freiburg
Vacuum Concentrator RC10.22	Jouan, Unterhaching
Vortex-Genie 2	Scientific Industries, Bohemia NY, USA
7900HT-RealTime-PCR System	Applied Biosystems, Foster City CA, USA

Table 8.3: Enzymes, proteins, and size markers.

Enzyme/Protein/Size marker	**Manufacturer**
Albumin Fraktion V	Roth, Karlsruhe
AvrII (4 u/μl)	New England Biolabs, Frankfurt a.M.
BfrBI/Mph1103I (NsiI, 10 u/μl)	Fermentas, St. Leon-Rot
BseRI (4 u/μl)	New England Biolabs, Frankfurt a.M.
Calf Intestine Alkaline Phosphatase (CIAP, 1 u/μl)	Fermentas, St. Leon-Rot
Calf Intestine Alkaline Phosphatase (CIP, 10 u/μl)	New England Biolabs, Frankfurt a.M.
Deoxyribonuclease I (DNase I, 1 u/μl)	Fermentas, St. Leon-Rot
DpnI (20 u/μl)	New England Biolabs, Frankfurt a.M.
ECL Anti-Mouse IgG (sheep), HPR-conjugated	GE Healthcare, München
ECL Anti-Rabbit IgG (donkey), HPR-conjugated	GE Healthcare, München
GeneRuler 1kb DNA Ladder	Fermentas, St. Leon-Rot
Hfq (purified *S. typhimurium* protein)	Sittka *et al.*, 2007
Lysozyme	Roth, Karlsruhe
Monoclonal Anti-FLAG M2 Antibody (mouse)	Sigma-Aldrich, Taufkirchen
Monoclonal Anti-GFP antibodies	Roche, Mannheim
(Clones 7.1, 13.1; mouse)	
NheI (10 u/μl)	Fermentas, St. Leon-Rot
pUC Mix Marker, 8	Fermentas, St. Leon-Rot
pUC19 DNA/*MspI* (HpaII) Marker, 23	Fermentas, St. Leon-Rot
Pfu DNA Polymerase (2.5 u/μl)	Fermentas, St. Leon-Rot
Phusion High-Fidelity DNA Polymerase (2 u/μl)	New England Biolabs, Frankfurt a.M.
Polyclonal Anti-GroEL antiserum (rabbit),	Sigma-Aldrich, Taufkirchen
Peroxidase conjugated	
Prestained Protein Marker Broad Range	New England Biolabs, Frankfurt a.M.
Ribonuclease H (RNase H, 5 u/μl)	New England Biolabs, Frankfurt a.M.
Ribonuclease T1 (RNase T1, 1 u/μl)	Ambion, Austin TX, USA
Ribonuclease T2 (RNase T2, 20 u/μl)	Invitrogen, Karlsruhe
Ribonuclease III (RNase III, 1.3 u/μl)	New England Biolabs, Frankfurt a.M.
RNA ladder Low Range, High Range	Fermentas, St. Leon-Rot
SpeI (10 u/μl)	New England Biolabs, Frankfurt a.M.
Shrimp Alkaline Phosphatase (SAP, 1 u/μl)	Fermentas, St. Leon-Rot
StratasScript Reverse Transcriptase	Stratagene, Cedar Creek TX, USA
SUPERase-In RNase Inhibitor (20 u/μl)	Ambion, Austin TX, USA
T4 DNA Ligase (5 Weiss u/μl)	Fermentas, St. Leon-Rot
T4 Polynucleotide Kinase (PNK, 10 u/μl)	Fermentas, St. Leon-Rot
T4 Polynucleotide Kinase (PNK, 10 u/μl)	New England Biolabs, Frankfurt a.M
T4 RNA Ligase (20 u/μl)	New England Biolabs, Frankfurt a.M.
Taq DNA polymerase (5 u/μl)	New England Biolabs, Frankfurt a.M.
Tobacco Acid Pyrophosphatase (TAP, 10 u/μl)	Epicentre, Madison WI, USA
Terminator^{TM} 5'-Phosphate-Dependent Exonuclease	Epicentre, Madison WI, USA

continued on next page

Enzyme/Protein/Size marker	Manufacturer
(TEX, 10 u/μl)	
XbaI (10 u/μl)	Fermentas, St. Leon-Rot
XhoI (20 u/μl)	New England Biolabs, Frankfurt a.M.
α-OppA polyclonal antibody (rabbit)	K. Igarashi, Chiba University, Japan

Table 8.4: Commercially available systems.

System	Application	Manufacturer
CycleReader^{TM} DNA Sequencing Kit	Sequencing ladders	Fermentas, St. Leon-Rot
Gel Loading Buffer II	RNA sample loading buffer	Ambion, Austin TX, USA
GlycoBlue^{TM}	DNA/RNA precipitation	Ambion, Austin TX, USA
MAXIscript T7 Kit	*In vitro* transcription for RNA used in EMSA, structure probing	Ambion, Austin TX, USA
MEGAscript T7 Kit	*In vitro* transcription for riboprobe synthesis	Ambion, Austin TX, USA
NucleoBond PC100	Plasmid isolation (large-scale)	Macherey-Nagel, Düren
NucleoSpin Extract II	DNA purification	Macherey-Nagel, Düren
NucleoSpin Plasmid QuickPure	Plasmid isolation (small-scale)	Macherey-Nagel, Düren
PageBlue Protein Staining Solution	Protein staining	Fermentas, St. Leon-Rot
Puresystem	*In vitro* translation	Cosmo Bio Co., Tokyo, Japan
Protein Loading Buffer Pack	Protein sample loading buffer	Fermentas, St. Leon-Rot
QuantitectTM SYBR Green RT-PCR Kit	Real time PCR	Qiagen, Hilden
Rapid-hyb^{TM} Buffer	Northern blot hybridization	Amersham Biosciences, Freiburg
Rediprime II DNA Labeling System	DNA probe synthesis	GE Healthcare, München
Roti-Free	Western blot stripping	Roth, Karlsruhe
Roti-Hybriquick	Northern blot hybridization	Roth, Karlsruhe
Stains-All	RNA staining	Sigma-Aldrich, München
SuperScriptII Reverse Transcriptase	cDNA synthesis	Invitrogen, Karlsruhe
SuperScriptIII Reverse Transcriptase	cDNA synthesis	Invitrogen, Karlsruhe
SV Total RNA Isolation System	DNA-free total RNA isolation	Promega, Madison WI, USA
TOPO TA Cloning Kit	Cloning of Taq PCR products	Invitrogen, Karlsruhe
TRIzol Reagent	total RNA isolation	Invitrogen, Karlsruhe

continued on next page

System	Application	Manufacturer
Western Lightning Chemiluminescence Reagent	Western blot detection	Perkin Elmer, Weiterstadt
6×Sample Loading Buffer	DNA loading buffer	Fermentas, St. Leon-Rot

8.2. General Methods

This section describes general methods that were applied throughout the whole study. Standard methods which are not described in this section were performed as described in Sambrook & Russell (2001), or according to manufacturers' instructions.

8.2.1. Bacterial cell culture

All materials used throughout this study were autoclaved for 20 minutes (min) at 121°C and 1 bar before use. Where necessary, solutions were sterilized by filtration and glassware by heating to 180°C for a minimum of three hours (h), respectively.

8.2.1.1. Media

If not stated otherwise, bacteria were grown in Lennox-broth (L-broth) or on Lennox agar plates.

Lennox broth:
1 % (w/v) tryptone
0.5 % (w/v) yeast extract
85.6 mM sodium chloride

Lennox agar:
as L-broth but supplemented with 1.2 % (w/v) agar.

Growth was carried out at 37°C with an agitation of 220 rotations per minute (rpm) under normal aeration. Cultures were either inoculated from single colony grown overnight at 37°C or were diluted 1/100 into fresh medium from overnight cultures that were inoculated from freshly grown single colonies into 3 ml L-broth.
For inducible P_{BAD} promoters, cultures were supplemented with 0.2 % (w/v) L-arabinose.

The following antibiotica were used throughout this study:

Kanamycin: 50 μg/ml
Ampicillin: 50 or 100 μg/ml for low- or high-copy plasmids, respectively
Chloramphenicol: 20 or 30 μg/ml for low- or high-copy plasmids, respectively
Streptomycin: 90 μg/ml

8.2.1.2. Preparation of electrocompetent *Salmonella* cells

For preparation of electrocompetent *Salmonella*, cells were inoculated either from single colony or 1/100 from overnight cultures in fresh medium. The cultures were grown at 37°C, 220 rpm until the suspensions reached an OD_{600} of 0.5. Cells were washed three times in ice-cold sterile water, resuspended in 50 μl ice-cold sterile water, and subjected to electroporation (1.8 kV, 25 μF, and 200 Ω). Following transformation, cells were resuspended in 1 ml SOC medium prior to plating. Recovery was carried out for 60 min at 37°C, 220 rpm before selection on L-broth agar plates with the appropriate antibiotics.

SOC medium
2 % (w/v) tryptone
0.5 % yeast extract
85.6 mM NaCl
2.5 mM KCl
10 mM $MgCl_2$
20 mM glucose

8.2.1.3. Transformation of chemically competent *E. coli* cells

1 μl of a ligation reaction or 0.5μl plasmid (concentration $\approx$ 10-100 ng/μl) was mixed with 20 μl of chemically competent *E. coli* TOP10 or TOP10 F', respectively (Invitrogen). After pre-incubation for 30 min on ice, cells were subjected to a heat shock for 30 seconds (s) at 42°C. Cells were chilled for 1 min on ice and resuspended in 100 μl SOC medium. Recovery was carried out for 60 min at 37°C, 220 rpm before selection on L-broth agar plates with the appropriate antibiotics.

8.2.1.4. Growth curves

For growth curve determination, 30 ml L-broth, supplemented with the appropriate antibiotic, were inoculated from overnight culture to an optical density OD_{600} of 0.04. Growth was carried out at 37°C, 220 rpm. The OD_{600} was measured in time intervals of 45 min over a time period of 630 min.

8.2.2. Mutant construction in *S. typhimurium* and *E. coli*

8.2.2.1. One-step inactivation of chromosomal genes

Chromosomal mutagenesis of *S. typhimurium* SL1344 followed the procedure described by Datsenko & Wanner (2000) for *E. coli* with a few modifications. The wild-type *Salmonella* strain carrying plasmid pKD46 was grown in L-broth supplemented with ampicillin and 0.2 % L-arabinose at 28°C to an OD_{600} of 0.5 (25 ml culture per transformation). Cells were collected by centrifugation (20 min, 4,000 rpm), washed three times with ice-cold water, and dissolved in 50 μl ice-cold water. Polymerase chain reaction (PCR) products of marker genes, *e. g.*, chloramphenicol or kanamycin resistance cassettes, (50 μl standard reactions) were DpnI-treated for 3 h at 37°C, and purified on agarose gels, followed by purification on Macherey-Nagel spin columns (NucleoSpin

Extract II). One-fifth of the 25 μl column eluate (in water) was used for transformation. 50 μl of competent cells were mixed with the purified PCR product in a chilled cuvette (0.1-cm electrode gap) and electroporated (18 kV/cm). Subsequently, 1 ml of pre-warmed SOC medium was added, and cells were recovered by incubation for 1 h at 37°C before selection on L-broth agar plates with the appropriate antibiotics. All mutations were moved to a fresh SL1344 background by phage P22 transduction.

8.2.2.2. Resistance removal following chromosomal one-step inactivation

For removal of resistance genes of mutant *Salmonella* the strain was transformed with the FLP recombinase-expressing, temperature sensitive helper plasmid pCP20 (Datsenko & Wanner, 2000) and transformants were selected at 28°C on plates containing ampicillin. Transformants were struck on three fresh plates, one containing no antibiotic, one containing ampicillin (as a control for removal of the pCP20 helper plasmid), and one containing chloramphenicol or kanamycin (as a control for removal of the chromosomal resistance cassette). Transformants grown overnight exclusively on plate containing no antibiotic were used for further experiments.

8.2.2.3. P22 transduction

P22 lysates were prepared from soft agar plate lysates of donor strains using P22 phage HT/105-1 by standard procedure (Sternberg & Maurer, 1991). 100 μl of an overnight culture of the donor strain were mixed with 3 ml TOP agar and poured on a pre-warmed L-broth plate.

TOP agar:
1 % (w/v) tryptone
1 % (w/v) agar
10 mM $MgSO_4$
5 mM $CaCl_2$
86 mM NaCl

100 μl of a P22 phage lysate were spread on the TOP agar surface followed by overnight incubation at 37°C. TOP agar was collected from the plate and resuspended in 5 ml L-broth containing 10 mM magnesium sulfate and 5 mM calcium chloride. Upon addition of 400 μl chloroform, the suspension was vigorously vortexed and incubated overnight at 4°C. After centrifugation (10 min, 4,000 rpm) the supernatant was transferred into a glass tube and 400 μl chloroform were added to the phage lysate. Storage was performed at 4°C.

For transduction, 100 μl of a culture of the acceptor strain grown from single colony to an OD_{600} of 1 were mixed with 1, 10, and 100 μl of the phage lysate and incubated for 15 min at room temperature. To stop the transduction, 100 μl of a 20 mM EGTA solution were added to 100 μl of the mixture. The entire 200 μl sample was plated on pre-warmed L-plates containing the appropriate antibiotic. Incubation was carried out for up to 3 days at 37°C. Transformants were verified by PCR.

8.2.3. Nucleic acids techniques

The concentrations of all nucleic acid solutions (DNA as well as RNA) were determined by measurements using a NanoDrop machine. For purification of PCR products or plasmid mini-preps the NucleoSpin Extract II and the NucleoSpin Plasmid QuickPure kits, respectively, from Macherey-Nagel were used. The standard methods of *in vitro* amplification of DNA by PCR and the ligation of DNA fragments were carried out as described in Sambrook & Russell (2001).

8.2.3.1. Agarose gel electrophoresis

Agarose gels were used to separate DNA fragments. For gel preparation, agarose was dissolved in concentrations of 0.8 to 2 % (w/v) in 1×TAE buffer.

50×TAE buffer:
242 g Tris base
57.1 ml acetic acid
100 ml 0.5 M EDTA
Adjust pH to 8.5, add H_2O to a final volume of 1 l.

At a gel solution temperature of 50-60°C, ethidium bromide was added to a final concentration of 40 μg/ 100 ml. Prior to loading, five volumes of sample were mixed with one volume of 6×sample loading buffer. Gels were run in 1×TAE buffer at 100 V for about 30-60 min (according to fragment size).

8.2.3.2. Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide (PAA) gels were used to separate RNA fragments of different size. For gel preparation, 40 % PAA solution was used. Native and denaturing gels were run in the presence of 0.5× and 1×TBE buffer, respectively.

10×TBE buffer:
0.89 M Tris Base
0.89 M boric acid
20 mM EDTA pH 8.0
Add H_2O to a final volume of 1 l.

Denaturing PAGE: For denaturing gels (the native structure of RNA molecules is destroyed), the gels are supplemented with urea to a final concentration of 8.3 M for Northern blot gels or 7 M for sequencing gels, respectively.

5 % PAA gel solution:
125 ml 40 % PAA solution (19:1 acrylamide/bisacrylamide)
500 g urea
100 ml 10×TBE

Add H2O to a final volume of 1 l.

All gel equipment was cleaned with 70 % ethanol before use (glass plates, spacer, combs etc.). Polymerization was initiated by addition of 1/100 volume of ammonium persulfate (APS) and 1/1000 volume of N,N,N,N,-Tetramethylethylendiamin (TEMED). Prior to loading, the RNA samples were denatured for 5 min at 100°C in sample loading buffer and subsequently chilled on ice for 5 min.

2×RPA loading buffer:
98 % (v/v) Formamid
2 mM EDTA, pH 8.0
0.02 % (w/v) Xylene Cyanole
0.02 % (w/v) Bromophenol Blue

Gel runs were performed in the presence of 1×TBE at 300 V at room temperature for about 2 to 3 h (according to the size of the RNA species to detect).

Native PAGE:

Native polyacrylamide gel electrophoresis was performed in the presence of 0.5×TBE at 300 V. Native PAGE was used to analyse preformed RNA-protein complexes. To avoid heating, the gel apparatus was connected to a water cooling system. Complexes were loaded in native sample loading buffer.

5×native sample buffer
50 % glycerol
0.2 % (w/v) Bromphenol Blue
0.5×TBE buffer

8.2.4. Protein techniques

8.2.4.1. Preparation of whole cell protein fraction

After the appropriate incubation, bacteria samples were taken (a total amount of 0.5 to 1 OD_{600}). After centrifugation for 2 min at 13,000 rpm at 4°C the supernatant was discarded and the cell pellet resuspended in 1× sample loading buffer to a final concentration of 0.01 OD/ μl buffer. The sample was denatured by heating at 95°C for 5 min and was subsequently chilled on ice.

8.2.4.2. Preparation of periplasmic fractions

The periplasmic protein fraction was extracted following the cold osmotic shock procedure described by Neu & Heppel (1965). At the appropriate OD_{600}, cells were harvested by centrifugation for 30 min at 4000 rpm at 4°C. The pellet was resuspended at room temperature in ‘shock buffer’.

Shock buffer:
30 mM Tris-HCl pH 8.0
20 % sucrose

EDTA (pH 8.0) was added to a final concentration of 1 mM. Cells were incubated for 10 min at room temperature with occasional shaking. Subsequently, cells were collected by centrifugation (30 min, 4,000 rpm) at 4°C, and the pellet was resuspended in 10 ml ice-cold 5 mM $MgSO_4$. After incubation for 10 min with occasional shaking in an ice-water bath, the suspension was centrifuged as mentioned above. The supernatant was the cold osmotic shock-fluid. For denaturation, 4× sample loading buffer was added and samples heated for 5 min at 95°C.

8.2.4.3. One-dimensional SDS PAGE

For denaturing separation of proteins, samples were loaded on 10 % to 15 % SDS PAGE (according to the size of the proteins to be analysed). Gel solutions for the separation and the stacking gel were prepared as follows:

PAA gel for separation gel	10 %	12 %	15 %
1 M Tris base pH 8.8	3.75 ml	3.75 ml	3.75 ml
40 % PAA solution (37.5:1 acrylamide/bisacrylamide)	2.5 ml	3 ml	3.75 ml
H_2O	3.75 ml	3.25 ml	2.5 ml
10 % SDS	100 μl	100 μl	100 μl
10 % APS	75 μl	75 μl	75 μl
TEMED	7.5 μl	7.5 μl	7.5 μl

PAA gel for stacking gel	
1 M Tris base pH 6.8	1.25 ml
40 % PAA solution (37.5:1 acrylamide/bisacrylamide)	1 ml
H_2O	7.485 ml
10 % SDS	100 μl
10 % APS	150 μl
TEMED	15 μl

10× running buffer:
250 mM Tris
1.92 M glycine
1 % SDS

All gel equipment was cleaned with 70 % ethanol before use (glass plates, spacer, combs etc.). Polymerization was initiated by addition of 10 % APS and TEMED (see above) to the separation and stacking gel, respectively. Gels were run for 1 h at 80 V (stacking gel) and for 2-6 h at $\approx$ 150 V (according to gel size and molecular weight of proteins to be detected). Gels were stained using PageBlue staining solution.

8.2.4.4. Two-dimensional SDS PAGE

Sample preparation from *Salmonella* cultures, analysis by high-resolution two-dimensional electrophoresis, protein staining, and peptide mass fingerprinting were performed at the Max-Planck-Institute for Infection Biology (Berlin, Germany) protein analysis core facility[1] according to previously published standard protocols (Doherty *et al.*, 1998; Jungblut & Seifert, 1990; Klose & Kobalz, 1995).

8.2.4.5. Western blot

1× Transfer buffer:
25 mM Tris base
192 mM glycine
20 % methanol

$TBST_{20}$ buffer:
20 mM Tris base
150 mM NaCl
0.1 % Tween 20

Whole-cell protein samples corresponding to 0.01 or 0.05 OD_{600} culture volume were separated via SDS-PAGE as described above. PVDF membranes were activated by incubation in methanol (90 s), H_2O (5 min), and transfer buffer (5 min), consecutively. Gels were blotted either for 60 min at 100 V at 4°C in a cable tank blotter or for 2 h at 2 mA/ cm^2 membrane in a semi-dry blotter onto PVDF membrane in transfer buffer. After rinsing in $1\times TBST_{20}$ buffer, membranes were blocked for 1 h in 10 % dry milk in $TBST_{20}$. Hybridization was carried out as follows: appropriate primary antisera or antibodies were diluted in 3 % BSA, $TBST_{20}$ and blots hybridized for 1 h at room temperature under agitation, followed by five 6 min wash-steps in $TBST_{20}$. Subsequently, the blots were hybridized with the α-rabbit-horseradish peroxidase (HRP) or α-mouse-HRP conjugated secondary antibodies (1:5,000 in 3 % BSA in $TBST_{20}$) for 1 h at room temperature. The final wash steps were performed $6\times$ for 10 min in $TBST_{20}$. Blots were developed using Western Lightning Reagent in a Fuji LAS-3000 CCD campera.

8.2.5. RNA techniques

Throughout this study, three different methods for isolation of total RNA from *E. coli*, *S. typhimurium* or *H. pylori* were used. RNA was always kept on ice and stored at -20°C.

8.2.5.1. RNA preparation with TRIzol

Bacterial cultures corresponding to 4 OD_{600} were spun for 2 min at 11,000 rpm at 4°C. After discarding the supernatant, the bacterial pellet was dissolved in 1 ml TRIzol Reagent. The mixture was transferred to 2 ml Phase lock tubes (heavy) and upon addition of 400 μl chloroform, the samples were mixed by shaking and centrifuged for 12 min at 15 °C at 13,000 rpm. The supernatant

[1] `http://info.mpiib-berlin.mpg.de/jungblut/`

was transferred to a fresh reaction tube and the nucleic acids were precipitated by addition of 0.7 volumes of 2-propanol. Precipitation was carried out either overnight or at least for 1 h at -20°C. Nucleic acids were pelleted by centrifugation for 30 min at 13,000 rpm at 4°C. After a wash step with 350 μl of 75 % ethanol and additional centrifugation for 10 min, the supernatant was discarded, the pellets air-dried, and subsequently dissolved in H_2O or 2$\times$ RNA loading buffer.

8.2.5.2. RNA isolation using the SV40 Total RNA Isolation System (Promega)

RNA was isolated using the Promega SV total RNA purification kit as described at the Institute for Food Research (Norwich, UK)[2] website and in Kelly *et al.* (2004). Two OD_{600} of a bacterial culture were mixed with 0.2 volume of stop-mix (ethanol:phenol 95:5 v/v). After snap-freezing in liquid nitrogen the samples were subsequently thawed on ice. Bacteria were spun for 2 min at 13,000 rpm at 4°C. After resuspension in 100 μl H_2O containing 50 mg/ml lysozyme the samples were incubated for 4 min at room temperature. Upon addition of 75 μl of lysis reagent, the samples were mixed and 350 μl RNA dilution buffer was added. The samples were heated for 3 min at 70°C, followed by a 10 min centrifugation step at 13,000 rpm at room temperature. The supernatant was transferred to a new tube, mixed with 200 μl 95 % ethanol and the mixture loaded on a spin column provided with the kit. After centrifugation for 1 min at maximum speed the eluate was discarded and the column washed with 600 μl wash buffer. After an additional centrifugation step, 50 μl of a DNase mix (5 μl 90 mM $MnCl_2$, 40 μl DNase core buffer and 5 μl DNase I; all provided with the kit) were applied to the membrane and incubation carried out for 15 min at room temperature. Upon addition of 200 μl DNase stop mix the columns were centrifuged for 1 min at 13,000 rpm. Following two wash steps with 600 and 250 μl wash buffer, respectively, (the second centrifugation was carried out for 2 min) the column was transferred to a sterile Eppendorf tube and 100 μl RNase-free water was added. After incubation for 1 min at room temperature, RNA was eluted by centrifugation for 2 min at 13,000 rpm.

8.2.5.3. Isolation of total RNA by sucrose shock and hot phenol

RNA preparation using hot phenol was first described in Aiba *et al.* (1981). In this study, cells were lysed using a 'sucrose shock' method followed by hot phenol extraction of RNA as described in Mattatall & Sanderson (1996). 10 OD_{600} of bacterial culture were mixed with 0.2 volume of stop-mix (ethanol:phenol 95:5 v/v) and incubated for 10 min on ice. Bacteria were pelleted by centrifugation for 20 min at 4,000 rpm at 4°C. The pellet was snap frozen in liquid nitrogen and stored at -80°C until RNA preparation.

Following resuspension of the pellet in 12 ml extraction buffer, 10% SDS solution was added to a final concentration of 1%.

Extraction buffer:
10 mM sodium acetate, pH 4.8
0.15 M sucrose

Cells were vigorously vortexed in 50 ml Falcon tubes; upon addition of 13 ml preheated (65°C) phenol the samples were transferred to a 65°C water bath and incubated for 5 min with short inter-

[2] `www.ifr.ac.uk/safety/microarrays/protocols.html`

vals of vortexing. After 30 min centrifugation in Corex tubes at 8,000 rpm at 4°C the upper phase was transferred to a new 50 ml Falcon tube and a second phenol extraction (13 ml Phenol, vortexing at room temperature) was carried out. After centrifugation in a Corex tube (see above) the upper phase was transferred to a fresh Corex tube and 12 ml chloroform were added to the sample. After 30 min centrifugation at 8,000 rpm at 4°C the upper phase was transferred into a fresh Corex tube and RNA was precipitated by addition of 2.5 volumes of a Ethanol:sodium acetate mixture (30:1 v/v, pH 6.5). The nucleic acids were precipitated overnight at -20°C. The samples were collected by centrifugation for 30 min at 8,000 rpm at 4°C. After aspiration of the supernatant, the pellet was washed with 4 ml 75% ethanol. Following subsequent centrifugation for 10 min at 8,000 rpm and 4°C, the supernatant was discarded and the pellets air dried. Samples were dissolved in water and subjected to DNase I treatment (see below).

8.2.5.4. DNase I digestion

Following resuspension in H_2O, RNA samples were treated with 1 u of DNase I per μg of RNA for 30 min at 37°C in the presence of RNase inhibitor. Prior to addition of DNase I reaction buffer and enzymes, the RNA was denatured at 65°C for 8 min and subsequently cooled on ice for 5 min. After DNase I digestion, the RNA was isolated by phenol:chloroform:isoamylalcohol (P:C:I) extraction. One volume dissolved RNA was mixed with one volume P:C:I (25:24:1 v/v) in 2 ml Phase lock (heavy) tubes. Following mixing for 15 s by vigorous shaking, samples were spun for 15 min at 13,000 rpm at 15°C. The aqueous (upper) phase was mixed with 2.5 volumes of 30:1 (v/v) EtOH:3M sodium acetate (pH 6.5) mixture and RNA was precipitated overnight at -20°C. After centrifugation for 30 min at 13,000 rpm and 4°C, the supernatant was discarded and the pellet was washed with 350 μl 75 % ethanol. After additional centrifugation for 10 min at 13,000 rpm and 4°C, the supernatant was discarded and the pellet was air-dried. Finally the RNA pellet was resuspended in H_2O.

8.2.5.5. Generation of radioactively labelled DNA oligo nucleotides for RNA detection

For labelling of DNA oligonucleotides, 10 pmol of the oligonucleotide was incubated in a 10 μl reaction volume with 25 μCi of ^{32}P-γ-ATP in the presence of 1 u T4 polynucleotide kinase (PNK, New England Biolabs) for 1 h at 37°C. Unincorporated nucleotides were removed using MicroSpin^{TM} G-25 columns.

8.2.5.6. Northern blot

To detect mRNAs or sRNAs, 5 to 20 μg RNA were separated on 5-10% denaturing (8.3 M urea) PAA gels. After a 1 h transfer to Hybond-XL membranes in a tankblotter at 50 V and 4°C in the presence of 1× TBE (see Section 8.2.3.2), the RNA was cross-linked to the membrane on a UV-table (302 nm) for 4 min or with 120 kJ in a UV-crosslinker (Bio-Link BLX 254, λ=254 nm). After prehybridization for 1 h in 15 ml Rapid-hyb buffer or 20 ml Roti-Hybri-Quick buffer at 42°C, the radioactive labelled probe (2-5 pmol) was added. After a period of 1 to 12 h of hybridization at 42 °C the membrane was rinsed with 5× SCC, followed by three wash steps at 42°C with SSC (20 min 5× SCC, 15 min 1× SCC, and 15 min 0.5× SCC). All SSC buffers were supplemented with 0.1% SDS.

20× SSC buffer:
3 M sodium chloride
0.3 M sodium citrate

8.2.5.7. Rapid amplification of cDNA ends (5´ RACE)

Mapping of transcriptional start sites using 5' RACE followed the protocol described in Argaman *et al.* (2001). Primary transcripts in bacteria carry a 5' triphosphate, which can be hydrolyzed by tobacco acid pyrophosphatase (TAP) specifically between the α- and β-phosphate groups. Via the resulting 5' monophosphate, these RNAs can be subsequently ligated to the 3' hydroxyl group of an RNA oligonucleotide, followed by reverse transcription with a gene-specific deoxyoligonucleotide and subsequent PCR amplification using a 5'-adapter specific primer and a nested gene-specific primer. TAP treatment is expected to yield a specific or at least strongly enhanced signal for primary transcripts in the amplification step as compared to untreated RNA samples. However, 5' ends resulting from processing (retaining a 5'-monophosphate) will also be amplified and can be analyzed in parallel.

In detail, 12 μg of total RNA was adjusted to a volume of 87.5 μl by adding the required volume of H_2O. 10 μl of 10×TAP buffer and 0.5 μl SUPERase-In RNase inhibitor were added and samples split into two reactions of 49 μl each. Following treatment of one reaction with 10 units tobacco acid pyrophosphatase (TAP) for 30 min at 37°C, 300 pmol of RNA-linker A4 (encoding a BseRI restriction sit) was added to both reactions. Following organic extraction and ethanol precipitation, RNA was dissolved in 13.5 μl H_2O, denatured for 5 min at 90°C and chilled on ice for 5 min. The RNA-linker ligation was performed overnight at 17°C in presence of 40 units T4 RNA ligase, 1×RNA ligase buffer, 10% v/v DMSO (final concentration) and 20 units SUPERase-In RNase Inhibitor. Following P:C:I extraction and ethanol precipitation, 2 μg linker-ligated RNA was converted to cDNA using 100 pmol random hexamer primers and the Superscript III (200 units) reverse transcription kit in a 20 μl reaction. 10 min incubation at 25°C was carried out before addition of the reverse transcriptase, followed by four subsequent 15 min incubation steps at 42°C, 50°C, 55°C and 60°C. After heat inactivation of the reverse transcriptase for 5 min at 85°C, samples were treated with 1 unit RNaseH (New England Biolabs) at 37°C for 20 min.

1 μl of the cDNA samples was subsequently used as template in standard Taq polymerase PCR reactions using the sense primer JVO-0367, which anneals to the RNA-linker sequence, and an antisense primer that anneals within the gene of interest. (For GFP fusion cloning, the antisense primer binds to the N-terminal coding region of the gene of interest and carries an in-frame NheI site.) Following visualization on 3% agarose gels, PCR products enriched in TAP treated samples (indicating primary transcripts) were excised, purified, and sequenced after TOPO cloning.

8.2.5.8. *In vitro* transcription and 5' end labelling of RNA

In vitro transcription was performed using the MEGAscript T7 kit, followed by DNase I digestion (1 unit) for 15 min at 37°C. Following extraction with P:C:I (25:24:1 v/v), unincorporated nucleotides were removed from the aqueous phase using MicroSpin^{TM} G-25 columns. The RNA was precipitated from the eluate by addition of 2.5 volumes of 30:1 EtOH:3M sodium acetate (pH 6.5) mixture and incubation overnight at -20°C. After centrifugation for 30 min at 13,000 rpm and

4°C the supernatant was discarded and the pellet washed with 350 μl 75 % ethanol. After additional centrifugation for 10 min at 13,000 rpm and 4°C, the supernatant was again discarded and the pellet air-dried. Finally, the RNA pellet was resuspended in H_2O. RNA integrity was checked on a denaturing PAA gel and visualized with Stains-All.

20 pmol RNA was dephosphorylated with 10 u of calf intestine alkaline phosphatase (CIP) in a 20 μl reaction at 37°C for 1 h. Following phenol extraction, the RNA was precipitated overnight with ethanol:sodium acetate (30:1 v/v) and 20 μg GlycoBlue. The dephosphorylated RNA was 5' end-labelled with ^{32}P-γ-ATP (20 μCi), using 1 u of T4 polynucleotide kinase (PNK) for 30 min at 37°C in a 20 μl reaction. Unincorporated nucleotides were removed using MicroSpin^{TM} G-25 columns, followed by purification of the labelled RNA on a denaturing gel (6% PAA/ 8.3 M urea). Upon visualization of the labelled RNA by exposure on a phosphorimager, the RNA was cut from the gel and eluted with RNA elution buffer at 4°C overnight, followed by P:C:I extraction and ethanol precipitation as before.

RNA elution buffer:
0.1 M sodium acetate
0.1% SDS
10 mM EDTA, pH 8.0

8.3. Methods: Multiple targeting of ABC transporter mRNAs by GcvB sRNA

8.3.1. Bacterial strains and oligonucleotides

Bacterial strains are listed in Table 10.1 in the Appendix. The *gcvB* deletion strains of *Salmonella* (JVS-0236) and *E. coli* (JVS-6081) were constructed using the λ red protocol (Datsenko & Wanner, 2000) as described in Section 8.2.2.1 by replacing residues 17-176 with a kanamycin marker gene, PCR amplified with primer JVO-0133/-0134 and JVO-0131/-0132, respectively. Mutants were verified by PCR with primers JVO-0135/-0136 or JVO-0137/-0138, respectively. The *Salmonella* Δ*gcvB*/Δ*hfq* strain (JVS-0617) was constructed by P22 transduction of Δhfq::Cm^R (JVS-0255) into strain JVS-0236. Marker genes were removed with FLP recombinase (Datsenko & Wanner, 2000) as described in Section 8.2.2.2.

DNA oligonucleotides used for cloning, PCR amplification of T7 templates, for toeprinting assays, and as hybridization probes are listed in the Appendix in Table 10.3.

8.3.2. Media and growth conditions

Cells were grown aerobically at 37°C in Lennox broth or M9 minimal medium supplemented with 0.4 % glucose. When required, antibiotics were added at 92 μg/μl streptomycin, 100 μg/ml ampicillin, 50 μg/ml kanamycin, and 20 μg/ml chloramphenicol.

Table 8.5: Construction of GcvB RNA mutant plasmids.

Plasmid	Name	Template	Oligo 1	Oligo 2	Deletion of position
pJL03-15	p*gcvB*$_{\Delta R1}$	pTP05	JVO-0746	JVO-0745	66 - 89
pJL16-10	p*gcvB*$_{\Delta R2}$	pTP05	JVO-0895	JVO-0896	136 - 144
pJL05-16	p*gcvB*$_{5'\Delta}$	pTP05	JVO-0743	JVO-0744	1 - 91
pJL01-1	p*gcvB*$_{3'\Delta}$	pTP05	JVO-0619	JVO-0742	135 - 206 and region downstream of *gcvB* terminator
pJL22	p$_L$*gcvB*$_{\Delta R1}$	pTP09	JVO-0746	JVO-0745	66 - 89
pJL23	p$_L$*gcvB*$_{\Delta R2}$	pTP09	JVO-0895	JVO-0896	136 - 144
pJL29-4	p$_L$*gcvB*$_{5'\Delta}$	pTP09	JVO-0744	P_{LlacO}-C	1 - 91

8.3.3. Plasmids

Plasmids are listed in Table 10.2 in the Appendix. Details of their construction and insert sequences are given in Tables 8.5-8.7.

8.3.3.1. sRNA plasmid construction

Control plasmid pTP11 was obtained by replacing the ColE1 origin of control plasmid pJV300 with the p15A origin of pZA31-*luc* via SpeI/AvrII cloning. pTP11 expresses a $\approx$ 50 nt nonsense transcript derived from the *rrnB* terminator. The low copy control vector pTP24 was constructed by replacing the ColE1 origin of pJV968-1 with the p15A origin of pTP05 by XhoI/AvrII cloning; it carries a neutral, 1.5 kb internal *lacZ* fragment.

The ColE1-based plasmid, pTP02, expressing *Salmonella gcvB* from its own promoter, was obtained by cloning a *gcvB* PCR fragment (primers JVO-0137/-0138) into the XhoI/XbaI sites of plasmid pZE12-*luc*. JVO-0137 binds 292 nt upstream of the +1 site of *gcvB*; JVO-0138 binds 116 nt downstream of the *gcvB* terminator. Since the pTP02 is lethal in *Salmonella*, the ColE1 origin was replaced with p15A as described above for pTP11. This yielded plasmid pTP05 (p*gcvB*).

To express *gcvB* from the P_{LlacO} promoter (plasmid pJV846-11, ColE1 origin), the GcvB gene was amplified using primers JVO-0237 (binds to the +1 site of *gcvB*) and JVO-0138 and inserted into pZE12-luc by blunt-end/XbaI cloning as described in Urban & Vogel (2007). Subsequently, the ColE1 origin of pJV846-11 was replaced by p15A as above (see pTP11), yielding pTP09 (pP_L *gcvB*).

GcvB mutant plasmids were constructed via PCR amplification from the original plasmids pTP05 or pTP09 using Phusion Polymerase, DpnI digestion of template plasmid, and self-ligation of purified PCR products. Primers and templates used for each mutant plasmid are listed in Table 8.5. Sequences of the mutant inserts are given in Table 8.6.

The *gcvB*$_{3'\Delta T}$ mutant plasmid (pJL13-12) was constructed by ligation of XhoI digested PCR products of pJV752-1 using oligos JVO-0619/ pZE-B and of pTP05 using primers JVO-0892/pZE-A. This plasmid contained only part of the modified terminator. PCR amplification with primers JVO-0619/-0892 on plasmid template pJL13-12, followed by self-ligation of the PCR product, gave plasmid p*gcvB*$_{3'\Delta T}$ (pJL17-6).

Table 8.6: Inserts of GcvB mutant plasmids. Red letters indicate the *gcvB* wild-type sequence, deleted parts of the individual GcvB mutants are marked in blue, and the modified terminator/inserted nucleotides for the 3' deletion mutants are shown in green. Black bold letters indicate the XbaI site for cloning.

Name	**Plasmid**	**Insert from +1 to end of *gcvB* terminator**	**Positions deleted as compared to wild-type *gcvB***
p*gcvB* pP_L*gcvB*	pTP05 pTP09	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttgtataacaaacggttagtt ttccagagccaccatctctttcacgtcagtacgattgatctgctg tttgttgccattagcgtctttatacgaaatcataccggtatcgtt atcgg**TCTAGA**	none
p*gcvB*$_{\Delta R1}$ pP_L*gcvB*$_{\Delta R1}$	pJL03-15 pJL22	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttgtataacaaacggttagtt ttccagagccaccatctctttcacgtcagtacgattgatctgctg tttgttgccattagcgtctttatacgaaatcataccggtatcgtt atcgg**TCTAGA**	66 - 89
p*gcvB*$_{\Delta R2}$ pP_L*gcvB*$_{\Delta R2}$	pJL16-10 pJL23	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttgtataacaaacggttagtt ttccagagccaccatctctttcacgtcagtacgattgatctgctg tttgttgccattagcgtctttatacgaaatcataccggtatcgtt atcgg**TCTAGA**	136 - 144
p*gcvB*$_{5'\Delta}$ pP_L*gcvB*$_{5'\Delta}$	pJL05-16 pJL29-4	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttgtataacaaacggttagtt ttccagagccaccatctctttcacgtcagtacgattgatctgctg tttgttgccattagcgtctttatacgaaatcataccggtatcgtt atcgg**TCTAGA**	1 - 91
p*gcvB*$_{3'\Delta}$	pJL01-1	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttgtataacaaacggttagtt ttccagagccaccatctctttcacgtcagtacgattgatctgctg tttgttgccattagcgtctttatacgaaatcataccggtatcgtt atcggc**TCTAGA**	135 - 206 and region downstream of *gcvB* terminator
p*gcvB*$_{3'\Delta T}$	pJL17-6	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcggaaacgctaccgttt tttttc**TCTAGA**	135 - 206; modification in terminator starts at position 121

Table 8.7: Inserts of GFP-fusion plasmids. *Salmonella* gene sequences are capitalized, in which black letters correspond to 5' UTR parts and red letters to ORF parts. BfrBI sites of pXG-10 and NheI sites that were used for cloning are highlighted in bold in magenta and green, respectively. The C/A-rich element of *gltI* is marked in blue.

GFP fusion	Insert
dppA::gfp	ATGAGGGGCATTTTATGGAGAATCCGCACTGCAACTCAGTCGATTATGCGAACGGAATCC CCACCTCTCACTACTGACCTGACCAGGTAAAAAACAAAAAAGGCCGGGCGGTAAAAGCCT TTGCAAAGGGCAAAACAACATACATCACAATTGGAGCAGAAGAATGAGTATTTCCTTGAA GAAGTCAGGGATGgctagc
oppA::gfp	ATCGACGAAAGGCGATCGAACGAATCGTCAGAATAAATAAAGTCGGTGATAGCAAAAGCA GTGACAGACCTGGCAGTACACCACCAGTGCTGCACAGGAACCCTGACGGGATTAAACAGG CTGGTAAAAACCAGTAATTATAATGAGTGGAGTACAAACACAATGTCTAACATCACGAAA AAAAGTTTGATTGCAGCGGGAATACTCACTGCtagc
livJ::gfp	GAGTATGCTGCTAAAGCACGGGTAGCTAGCCAATAATCGAAATAAAGTGCTGAACAATAA CACCACAACACACGTAACAACCAGAATAATGGGGATTATCAGGATGAATATGAAGGGTAA AACGTTATTGGCAGGATGTATCGCtagc
livK::gfp	atgcatATCTATAGCGAAAAGCAGAATATTATCTTTTCTTAATAGACTGAAAAATAGAGA TTTTAATCTTATTATGCTTTAAATGCTGCGCTAACTCATTAATGAGTCAGTAAAAAGCGC ACCATTTATAAAAAGTACAGTCTGCTTTTTAACCAGCAAAAAACAAAACATATAACATCA CGAATGGGGATACAGGCACATGAAACGGAAAGCGAAAACAATAATCGCAGGGATTGTTGC tagc
argT::gfp	AGGACAATATTGCAACGTTTTATTAACAAATTTAACGTCGAATCGTTTTGCTGACGTGAA AATGGCATAAGACCTGCATGAAAAAGTCTGCAAACACACAACGCCACGTAAAACATAAGA AAATGACGCCACTTGAGGGGTATGTATGAAGAAGACCGTTCTCGCTTTGTCTTTGCTGAT AGGTCTGGGCGCtagc
STM4351::*gfp*	atgcatATCAGAATAGCACCCTGCGCCAAAAAAAGAATAGCACGGTGACCACAACATCCA ATTGATATCAGGGATCAAGATGAAAAAAAAACTTATTGTCATGCTGTTAGCGCtagc
gltI::gfp	ATAACACTGCACGCGCAAGTTGCAGGCAATAACAACATCACAATAGCTATCAATGCGTCG ACGGCGCAGATGATAAAGGAGTTGGATATGCAATTACGTAAGCTAACCACAGCAATGCTG GTCATGGGACTGTCTGCtagc
gltI$_{\Delta CA}$*::gfp*	ATAACACTGCACGCGCAAGCTATCAATGCGTCGACGGCGCAGATGATAAAGGAGTTGGAT ATGCAATTACGTAAGCTAACCACAGCAATGCTGGTCATGGGACTGTCTGCtagc
ompR::gfp	atgcatGCTTTTTTAAGAATACACGCTTACAAATTGTTGCGAACCTTTGGGAGTACAAAC AATGCAAGAGAACTACAAGATTCTGGTGGTCGATGACGACATGCGCCTGCGTGCGCTGCT GGAACGTTATCTCACCGAACAAGGCTTCCAGGTTCGAAGCGTCGCTagc
ompR$_{CA}$*::gfp*	atgcatGCTTTTTTAAGAATAGTTGCAGGCAATAACAACATCACAATACACGCTTACAAA TTGTTGCGAACCTTTGGGAGTACAAACAATGCAAGAGAACTACAAGATTCTGGTGGTCGA TGACGACATGCGCCTGCGTGCGCTGCTGGAACGTTATCTCACCGAACAAGGCTTCCAGGT TCGAAGCGTCGCTagc

8.3.3.2. Fusion plasmid construction

Translational GFP fusions to GcvB target mRNAs were constructed as described in Urban & Vogel (2007). For *dppA*, *oppA*, *gltI*, *livJ*, and *argT* from *Salmonella* SL1344, promoters were confirmed or mapped by 5' RACE as described in Section 8.2.5.7; BseRI/NheI digested 5'RACE cDNA fragments were cloned into the BsgI/NheI-digested fusion plasmid pXG-20. The *livK* and STM4351 fusions were cloned in vector pXG-10. The insert sequences of all GFP fusions are given in Table 8.7. Details for GFP plasmid construction are listed in Table 8.8.

To construct fusion *gltI*$_{\Delta CA}$*::gfp* (pJL45-3), bp -71 to -44 relative to *gltI* AUG were deleted from the *gltI::gfp* fusion plasmid by PCR amplification of pJL24-1 using oligos JVO-1973/-1974 and Phusion Polymerase, DpnI digestion, and self-ligation of purified PCR product. The deleted 27 bp C/A-rich element was inserted in the *ompR::gfp* fusion (pJU-63) at position -42 according to the

Table 8.8: Details of GFP-fusion plasmids construction. .

Target gene	Plasmid trivial name	Plasmid original name	Backbone	Oligos	Insert digested with	5'RACE end/TAP specific	Ecocyc promoter[a]	Insert 5' end [bp]	Insert 3' end [bp]	Total insert size [bp]	GFP fused to aa	Signal sequence [aa][b]	Fluorescence on plate	Comment
dppA	*dppA::gfp*	pJL19-1	pXG-20	JVO-0367 × JVO-0424	BseRI/ NheI	yes/+	-165	-163	30	193	10	28	+	
oppA	*oppA::gfp*	pJL18-1	pXG-20	JVO-0367 × JVO-0656	BseRI/ NheI	yes/-	-511 (-266, -171)[c]	-162	50	212	17	26	++	
gltI [d]	*gltI::gfp*	pJL24-1	pXG-20	JVO-0367 × JVO-0427	BseRI/ NheI	yes/-	-149	-87	50	137	17	22	+++	ATG wrongly annotated in *Salmonella*
livJ [d]	*livJ::gfp*	pJL20-1	pXG-20	JVO-0367 × JVO-0728	BseRI/ NheI	yes/-	-104	-103	41	144	14	23	+++	ATG wrongly annotated in *Salmonella*
livK	*livK::gfp*	pJL31-24	pXG-10	JVO-1271 × JVO-0800	BfrBI/ NheI	no product	-195/ -170	-193	41	234	14	23	-	
argT	*argT::gfp*	pJL27-2	pXG-20	JVO-0367 × JVO-0796	BseRI/ NheI	yes/+	-60	-145	47	192	16	22	-	
STM4351	STM4351*::gfp*	pTP28	pXG-10	JVO-0731 × JVO-0732	BfrBI/ NheI	no[e]/-	-	-73	32	105	11	17	-	

[a] Promoter in *E. coli* based on EcoCyc `www.ecocyc.org` annotation.

[b] Signal peptide predicted with SignalP 3.0 server (`http://www.cbs.dtu.dk/services/SignalP/`); (Bendtsen *et al.*, 2004).

[c] Promoters 2 and 3 of *oppA* from (Igarashi *et al.*, 1997).

[d] Positions and amino acid numbers are given according to the start codon that was determined by toeprinting (*gltI*) and sequence alignment with homologous genes from other bacteria (*livJ*).

[e] The 5' end of STM4351 was mapped at position -65.

ompR AUG to yield *ompR*$_{CA}$*::gfp* (pJL50-11). This plasmid was constructed by PCR amplification of pJU-63 using oligos JVO-2154/-2155, followed by self-ligation of the purified PCR-product.

8.3.4. RNA and protein detection

RNA preparation and Northern analysis followed previously published protocols (Urban & Vogel, 2007) and as described in Section 8.2.5. GcvB RNAs were detected with 5' end-labelled oligos (see figure legends), and 5S rRNA or *gfp* fusion mRNAs with oligos JVO-0322 or JVO-0155, respectively.

OppA protein was detected on Western Blot using a polyclonal OppA antibody kindly provided by K. Igarashi (Chiba University, Japan) according to the protocol described in Section 8.2.4 and Sittka *et al.* (2007). GFP fusion and GroEL proteins were detected as described in Urban & Vogel (2007). Periplasmic proteins were prepared from *Salmonella* cultured to an OD_{600} of 2 as previously published in Sittka *et al.* (2007) and described in Section 8.2.4.2, and analysed by high-resolution 2D electrophoresis, protein staining, and peptide mass fingerprinting at the Max-Planck-Institute for Infection Biology (Berlin, Germany) protein analysis core facility[3] according to previously published standard protocols.

8.3.5. Colony fluorescence imaging

E. coli carrying *gfp* fusion plasmids were grown overnight on LB plates supplemented with the appropriate antibiotics. Colonies were photographed using a CCD camera after 2 s excitation at 460 nm with a 510 nm emission filter in a FUJI LAS-3000 image analyser.

8.3.6. T7 transcription, purification and 5' end labelling of RNA

DNA templates carrying a T7 promoter sequence for *in vitro* transcription were generated by PCR. Primers and sequences of the T7 transcripts are given in Tables 8.9 and 8.10. T7 templates of *gfp* fusion mRNAs were amplified from plasmids using a sense primer that adds a T7 promoter to the +1 site of the 5'UTR, and antisense oligo pZE-T1, which binds 122 nt downstream of the *gfp* stop codon. These transcripts end with the *rrnB* terminator of the fusion plasmids. RNA was *in vitro* transcribed and quality-checked as previously published in Sittka *et al.* (2007) and described in Section 8.2.5.8. The protocol for 5' end labelling of RNA is published in Papenfort *et al.* (2006) and also described in Section 8.2.5.8.

8.3.7. Gel mobility shift assays

GcvB/*dppA* leader and GcvB/*oppA* leader binding assays were performed in 1× structure buffer (provided with RNase T1) in a total reaction volume of 10 μl as follows. 5' end-labelled GcvB RNA ($\approx$ 5 nM final concentration in binding reaction) and 1 μg of yeast RNA (Ambion) were incubated in the presence of unlabelled *dppA* or *oppA* leader (final concentrations are given in the figure legends) at 37°C for 15 min. Prior to gel loading, the binding reactions were mixed with 3 μl

[3] `http://info.mpiib-berlin.mpg.de/jungblut/`

Table 8.9: Details of RNAs used for *in vitro* work. .

***In vitro* synthesized RNA fragments**	**Oligos**[a]	**Template**	**5' part to ATG [nt]**[b]	**3' part from ATG [nt]**[b]	**Size [nt]**
GcvB WT	JVO-0941/ JVO-0942	pTP05	-	-	201
GcvB ΔR1	JVO-0941/ JVO-0942	pJL3-15/ pJL-22	-	-	177
GcvB ΔR2	JVO-0941/ JVO-0942	pJL16-10/ pJL-23	-	-	192
GcvB 3'Δ	JVO-0941/ JVO-0742	pTP05	-	-	134
MicA	JVO-0937/ JVO-0938	pKP6-21	-	-	73
dppA leader	JVO-1034/ JVO-1035	genomic DNA SL1344	- 163	+ 72	236
oppA leader	JVO-1037/ JVO-1038	genomic DNA SL1344	- 162	+ 57	219
gltI leader	JVO-1039/ JVO-1040	genomic DNA SL1344	- 87	+ 74	161
livJ leader	JVO-1065/ JVO-1066	genomic DNA SL1344	-103 (- 109)	+ 66 (+ 60)	169
livK leader	JVO-1063/ JVO-1064	genomic DNA SL1344	- 193	+ 58	251
STM4351 leader	JVO-1067/ JVO-1068	genomic DNA SL1344	- 73	+ 85	158
argT leader	JVO-1060/ JVO-1061	genomic DNA SL1344	- 145	+ 58	236
ompA leader	JVO-1768/ JVO-1769	genomic DNA SL1344	- 133	+ 38	171
gfp leader	JVO-1048/ JVO-1049	pXG-1 (pJV859-8)	- 47	+ 70	117
gltI::gfp	JVO-1039/ pZE-T1	pJL24-1	- 87	+ 887	974
$gltI_{\Delta CA}$*::gfp*	JVO-1039/ pZE-T1	pJL45-3	- 60	+ 887	947
$gltI_{M2}$*::gfp*	JVO-1039/ pZE-T1	pJL56-2	- 87	+ 887	974
ompR::gfp	JVO-2233/ pZE-T1	pJU-63	- 55	+ 941	996
$ompR_{CA}$*::gfp*	JVO-2234/ pZE-T1	pJL50-11	- 82	+ 941	1023
cycA leader	JVO-1274/ JVO-1042	genomic DNA SL1344	- 79	+ 82	161
cycA_10th::gfp leader	JVO-1274/ JVO-1976	pJL83-2	- 79	+ 92	171

[a] Oligos that were used for amplification of T7 template.

[b] Only for mRNAs.

of native loading buffer and electrophoresed on native 6% polyacrylamide gels in 0.5× TBE buffer at 300 V at 4°C for 3 h (see Section 8.2.3.2). Gels were dried, and analysed on X-ray films. Gel shifts with labelled *dppA/oppA* leader and unlabelled GcvB RNA were done in the same way.

10× Structure buffer:
100 mM Tris, pH 7.0
1 M KCl
100 mM MgCl2

Table 8.10: Sequences of T7 transcripts used for *in vitro* work (cf. Table 8.9).

Red letters indicate ORFs in the T7 transcripts and start codons are shown in bold. Lower-case letters correspond to the *gfp* coding region and the *gfp* stop-codon UAA is shown in black bold letters. NheI (GCTAGC) and XbaI (TCTAGA) sites of the template plasmids are highlighted in green. The C/A-rich element of *gltI* is marked in magenta, whereas dark blue letters indicate the plasmid-borne *rrnB* terminator.

T7 RNA	Sequence
GcvB WT	ACUUCCUGAGCCGGAACGAAAAGUUUUAUCGGAAUGCGUGUUCUGAUGGGCUUUUGGCUUACGGUUGUGA UGUUGUGUUGUUGUGUUUGCAAUUGGUCUGCGAUUCAGACCACGGUAGCGAGACUACCCUUUUUCACUUC CUGUACAUUUACCCUGUCUGUCCAUAGUGAUUAAUGUAGCACCGCCAUAUUGCGGUGCUUU
GcvBΔR1	ACUUCCUGAGCCGGAACGAAAAGUUUUAUCGGAAUGCGUGUUCUGAUGGGCUUUUGGCUUACGGUCAAUU GGUCUGCGAUUCAGACCACGGUAGCGAGACUACCCUUUUUCACUUCCUGUACAUUUACCCUGUCUGUCCA UAGUGAUUAAUGUAGCACCGCCAUAUUGCGGUGCUUU
GcvBΔR2	ACUUCCUGAGCCGGAACGAAAAGUUUUAUCGGAAUGCGUGUUCUGAUGGGCUUUUGGCUUACGGUUGUGA UGUUGUGUUGUUGUGUUUGCAAUUGGUCUGCGAUUCAGACCACGGUAGCGAGACUACCCUUUUUCACAUU UACCCUGUCUGUCCAUAGUGAUUAAUGUAGCACCGCCAUAUUGCGGUGCUUU
GcvBΔ3'end	ACUUCCUGAGCCGGAACGAAAAGUUUUAUCGGAAUGCGUGUUCUGAUGGGCUUUUGGCUUACGGUUGUGA UGUUGUGUUGUUGUGUUUGCAAUUGGUCUGCGAUUCAGACCACGGUAGCGAGACUACCCUUUUU
MicA	GAAAGACGCGCAUUUGUUAUCAUCAUCCCUGUUUUCAGCGAUGAAAUUUUGGCCACUCCGUGAGUGGCCU UUU
dppA leader	AUGAGGGGCAUUUUAUGGAGAAUCCGCACUGCAACUCAGUCGAUUAUGCGAACGGAAUCCCCACCUCUCA CUACUGACCUGACCAGGUAAAAAACAAAAAAGGCCGGGCGGUAAAAGCCUUUGCAAAGGGCAAAACAACA UACAUCACAAUUGGAGCAGAAGA**AUG**AGUAUUUCCUUGAAGAAGUCAGGGAUGCUGAAGCUUGGUUUGAG CCUGGUGGCCAUGACCGUUGCAGCAA
oppA leader	AUCGACGAAAGGCGAUCGAACGAAUCGUCAGAAUAAAUAAAGUCGGUGAUAGCAAAAGCAGUGACAGACC UGGCAGUACACCACCAGUGCUGCACAGGAACCCUGACGGGAUUAAACAGGCUGGUAAAAACCAGUAAUUA UAAUGAGUGGAGUACAAACACA**AUG**UCUAACAUCACGAAAAAAAGUUUGAUUGCAGCGGGAAUACUCACU GCGCUCAUC
gltI leader	AUAACACUGCACGCGCAAGUUGCAGGCAAUAACAACAUCACAAUAGCUAUCAAUGCGUCGACGGCGCAGA UGAUAAAGGAGUUGGAU**AUG**CAAUUACGUAAGCUAACCACAGCAAUGCUGGUCAUGGGACUGUCUGCGGG CCUUGCGCACGCAGAAGAUGG
livJ leader	GAGUAUGCUGCUAAAGCACGGGUAGCUAGCCAAUAAUCGAAAUAAAGUGCUGAACAAUAACACCACAACA CACGUAACAACCAGAAUAAUGGGGAUUAUCAGG**AUG**AAU**AUG**AAGGGUAAAACGUUAUUGGCAGGAUGUA UCGCCCUGUCAUUAAGCCAUAUGGCAUUC
livK leader	AUCUAUAGCGAAAAGCAGAAUAUUAUCUUUUCUUAAUAGACUGAAAAAUAGAGAUUUUAAUCUUAUUAUG CUUUAAAUGCUGCGCUAACUCAUUAAUGAGUCAGUAAAAAGCGCACCAUUUAUAAAAAGUACAGUCUGCU UUUUAACCAGCAAAAAACAAAACAUAUAACAUCACGAAUGGGGAUACAGGCAC**AUG**AAACGGAAAGCGAA AACAAUAAUCGCAGGGAUUGUUGCAUUAGCAGUCUCGCAGG
STM4351 leader	AUCAGAAUAGCACCCUGCGCCAAAAAAAGAAUAGCACGGUGACCACAACAUCCAAUUGAUAUCAGGGAUC AAG**AUG**AAAAAAAAACUUAUUGUCAUGCUGUUAGCCAGCCUCUCCGUUCACGCCGCUUCCGUUUCCGCCA GAACAUUACAUUUUGGCA

continued on next page

T7 RNA	**Sequence**
argT leader	AGGACAAUAUUGCAACGUUUUAUUAACAAAUUUAACGUCGAAUCGUUUUGCUGACGUGAAAAUGGCAUAA GACCUGCAUGAAAAAGUCUGCAAACACACAACGCCACGUAAAACAUAAGAAAAUGACGCCACUUGAGGGG UAUGU**AUG**AAGAAGACCGUUCUCGCUUUGUCUUUGCUGAUAGGUCUGGGCGCGACGGCGGCCAGUUACGC CGCGCUCCCGCAAACGGUUCGUAUUG
ompA leader	GCCAGGGGTGCTCAGCATAAGCCGTAGATATCGGTAGAGTAACTATTGAGCAGATCCCCCGGTGAAGGAT TTAACCGTGTTATCTCGTTGGAGATATTCATGGCGTATTTTGGATGATAACGAGGCGCAAAAA**AUG**AAAA AGACAGCTATCGCGATTGCAGTGGCACTGGC
gfp leader	AUCAGCAGGACGCACUGACCGAAUUCAUUAAAGAGGAGAAAGGUACC**AUG**GCUAGCAAAGGAGAAGAACU UUUCACUGGAGUUGUCCCAAUUCUUGUUGAAUUAGAUGGUGAUGUUA
gltI::gfp	AUAACACUGCACGCGCAAGUUGCAGGCAAUAACAACAUCACAAUAGCUAUCAAUGCGUCGACGGCGCAGA UGAUAAAGGAGUUGGAUAUGCAAUUACGUAAGCUAACCACAGCAAUGCUGGUCAUGGGACUGUCUGcuag caaaggagaagaacuuuucacuggaguugucccaauucuuguugaauuagauggugauguuaaugggcac aaauuuucugucaguggagagggugaaggugaugcuacauacggaaagcuuacccuuaaauuuauuugca cuacuggaaaacuaccuguuccauggccaacacuugucacuacuuugaccuaugguguucaaugcuuuuc ccguuauccggaucauaugaaacggcaugacuuuuucaagagugccaugcccgaagguuauguacaggaa cgcacuauaucuuucaaagaugacgggaacuacaagacgcgugcugaagucaaguuugaaggugauaccc uuguuaaucguaucgaguuaaaagguauugauuuuaaagaagauggaaacauucucggacacaaacucga guacaacuauaacucacacaauguauacaucacggcagacaaacaaaagaauggaaucaaagcuaacuuc aaaauucgccacaacauugaagauggauccguucaacuagcagaccauuaucaacaaaauacuccaauug gcgauggcccuguccuuuuaccagacaaccauuaccugucgacacaaucugcccuuucgaaagaucccaa cgaaaagcgugaccacaugguccuucuugaguuuguaacugcugcugggauuacacauggcauggaugag cucuacaaa**TAA**ugaauucgagcauuuaaaucuagaggcaucaaauaaaacgaaaggcucagucgaaaga cugggccuuucguuuuaucuguuguuuugucgguaacgcucuccugaguaggacaaauccgccg
gltI$_{\Delta CA}$*::gfp*	AUAACACUGCACGCGCAAGCUAUCAAUGCGUCGACGGCGCAGAUGAUAAAGGAGUUGGAUAUGCAAUUAC GUAAGCUAACCACAGCAAUGCUGGUCAUGGGACUGUCUGcuagcaaaggagaAgaacuuuucacuggagu ugucccaauucuuguugaauuagauggugauguuaaugggcacaaauuuuucugucaguggagagggugaa ggugaugcuacauacggaaagcuuacccuuaaauuuauuugcacuacuggaaaacuaccuguuccauggc caacacuugucacuacuuugaccuaugguguucaaugcuuuucccguuauccggaucauaugaaacggca ugacuuuuucaagagugccaugcccgaagguuauguacaggaacgcacuauaucuuucaaagaugacggg aacuacaagacgcgugcugaagucaaguuugaaggugauacccuuguuaaucguaucgaguuaaaaggua uugauuuuaaagaagauggaaacauucucggacacaaacucgaguacaacuauaacucacacaauguaua caucacggcagacaaacaaaagaauggaaucaaagcuaacuucaaaauucgccacaacauugaagaugga uccguucaacuagcagaccauuaucaacaaaauacuccaauuggcgauggcccuguccuuuuaccagaca accauuaccugucgacacaaucugcccuuucgaaagaucccaacgaaaagcgugaccacaugguccuucu ugaguuuguaacugcugcugggauuacacauggcauggaugagcucuacaaa**TAA**ugaauucgagcauuu aaaucuagaggcaucaaauaaaacgaaaggcucagucgaaagacugggccuuucguuuuaucuguuguuu gucggugaacgcucuccugaguaggacaaauccgccg
gltI$_{M2}$*::gfp*	AUAACACUGCACGCGCAAGUUGCAGGCAAUAACAACAUCACAAUAGCUAUGAAUCCGUCGACGGCGCAGA UGAUAAAGGAGUUGGAUAUGCAAUUACGUAAGCUAACCACAGCAAUGCUGGUCAUGGGACUGUCUGcuag caaaggagaagaacuuuucacuggaguugucccaauucuuguugaauuagauggugauguuaaugggcac aaauuuucugucaguggagagggugaaggugaugcuacauacggaaagcuuacccuuaaauuuauuugca cuacuggaaaacuaccuguuccauggccaacacuugucacuacuuugaccuaugguguucaaugcuuuuc ccguuauccggaucauaugaaacggcaugacuuuuucaagagugccaugcccgaagguuauguacaggaa cgcacuauaucuuucaaagaugacgggaacuacaagacgcgugcugaagucaaguuugaaggugauaccc uuguuaaucguaucgaguuaaaagguauugauuuuaaagaagauggaaacauucucggacacaaacucga guacaacuauaacucacacaauguauacaucacggcagacaaacaaaagaauggaaucaaagcuaacuuc aaaauucgccacaacauugaagauggauccguucaacuagcagaccauuaucaacaaaauacuccaauug gcgauggcccuguccuuuuaccagacaaccauuaccugucgacacaaucugcccuuucgaaagaucccaa cgaaaagcgugaccacaugguccuucuugaguuuguaacugcugcugggauuacacauggcauggaugag cucuacaaa**TAA**ugaauucgagcauuuaaaucuagaggcaucaaauaaaacgaaaggcucagucgaaaga cugggccuuucguuuuaucuguuguuuugucgguaacgcucuccugaguaggacaaauccgccg

continued on next page

T7 RNA	Sequence
ompR::gfp	GCUUUUUUAAGAAUACACGCUUACAAAUUGUUGCGAACCUUUGGGAGUACAAACAAUGCAAGAGAACUAC AAGAUUCUGGUGGUCGAUGACGACAUGCGCCUGCGUGCGCUGCUGGAACGUUAUCUCACCGAACAAGGCU UCCAGGUUCGAAGCGUCGCUagcaaaggagaagaacuuuucacuggaguugucccaauucuuguugaauu agauggugauguuaaaugggcacaaauuuucugucaguggagagggugaaggugaugcuacauacggaaag cuuacccuuaaauuuauuugcacuacuggaaaacuaccuguuccauggccaacacuugucacuacuuuga ccuaugguguucaaugcuuuucccguuauccggaucauaugaaacggcaugacuuuuucaagagugccau gcccgaaggauauguacaggaacgcacuauaucuuucaaagaugacgggaacuacaagacgcgugcugaa gucaaguuugaaggugauacccuuguuaaucguaucgaguuaaaaggauuugauuuuaaagaagauggaa acauucucggacacaaacucgaguacaacuauaacucacacaauguauacaucacggcagacaaacaaaa gaauggaaucaaagcuaacuucaaaauucgccacaacauugaagauggauccguucaacuagcagaccau uaucaacaaaauacuccaauuggcgauggcccuguccuuuuaccagacaaccauuaccugucgacacaau cugcccuuucgaaagaucccaacgaaaagcgugaccacaugguccuucuugaguuuguaacugcugcugg gauuacacauggcauggaugagcucuacaaaTAAugaauucgagcauuuaaaucuagaggcaucaaauaa aacgaaaggcucagucgaaagacugggccuuucguuuuaucuguuguuugucggugaacgcucuccugag uaggacaaauccgccg
ompR$_{CA}$::gfp	GCUUUUUUAAGAAUAGUUGCAGGCAAUAACAACAUCACAAUACACGCUUACAAAUUGUUGCGAACCUUUG GGAGUACAAACAAUGCAAGAGAACUACAAGAUUCUGGUGGUCGAUGACGACAUGCGCCUGCGUGCGCUGC UGGAACGUUAUCUCACCGAACAAGGCUUCCAGGUUCGAAGCGUCGCUagcaaaggagaagaacuuuucac uggaguugucccaauucuuguugaauuagauggugauguuaaaugggcacaaauuuucugucaguggagag ggugaaggugaugcuacauacggaaagcuuacccuuaaauuuauuugcacuacuggaaaacuaccuguuc cauggccaacacuugucacuacuuugaccuaugguguucaaugcuuuucccguuauccggaucauaugaa acggcaugacuuuuucaagagugccaugcccgaaggauauguacaggaacgcacuauaucuuucaaagau gacgggaacuacaagacgcgugcugaagucaaguuugaaggugauacccuuguuaaucguaucgaguuaa aagguauugauuuuaaagaagauggaaacauucucggacacaaacucgaguacaacuauaacucacacaa uguauacaucacggcagacaaacaaaagaauggaaucaaagcuaacuucaaaauucgccacaacauugaa gauggauccguucaacuagcagaccauuaucaacaaaauacuccaauuggcgauggcccuguccuuuuac cagacaaccauuaccugucgacacaaucugcccuuucgaaagaucccaacgaaaagcgugaccacaugguc ccuucuugaguuuguaacugcugcugggauuacacauggcauggaugagcucuacaaaTAAugaauucga gcauuuaaaucuagaggcaucaaauaaaacgaaaggcucagucgaaagacugggccuuucguuuuaucug uuguuugucggugaacgcucuccugaguaggacaaauccgccg
cycA leader	GUCCUGAUAACAGGAUCGUCGUAUCAUAGACCAAAGGCCGUAGAGCCCGCACAACACAGACAGGUACAGG AAGAAAAACAUGGUAGAUCAGGUAAAAGUCGCAGCCGACGAACAGGCUCCGGCUGAACAGUCGCUACGGC GCAAUCUUACAAACCGUCAUA
cycA_10th::gfp leader	GUCCUGAUAACAGGAUCGUCGUAUCAUAGACCAAAGGCCGUAGAGCCCGCACAACACAGACAGGUACAGG AAGAAAAACAUGGUAGAUCAGGUAAAAGUCGCAGCCGACgctagcaaaggagaagaacuuuucacuggag uugucccaauucuuguugaauuagaugguga

8.3.8. *In vitro* structure mapping and footprinting

Secondary structure probing and mapping of RNA complexes was conducted on 5'-end-labelled RNA (≈0.1 pmol) in 10 μl reactions. RNA was denatured for 1 min at 95°C and chilled on ice for 5 min, upon which 1 μg yeast RNA and 10×structure buffer were added. Concentrations of unlabelled sRNA/mRNA leader added to the reactions are given in the figure legends. Following incubation for 10 min at 37°C, 2 μl of a fresh solution of lead(II) acetate (25 mM) or 2 μl of RNase T1 (0.01 u/μl) or 2 μl RNase T2 (0.02 u/μl) were added and incubated for 2, 3 or 5 min at 37°C, respectively. RNase III cleavage reactions contained 1 mM DTT and 1.3 unit enzyme, and were incubated for 6 min at 37 °C.

Reactions were stopped with 5 μl of 0.1 M EDTA, precipitated, and dissolved in Gel Loading Buffer II (Ambion), or by direct addition of 12 μl loading buffer on ice. RNase T1 ladders were obtained by incubating labelled RNA (≈0.2 pmol) in 1× sequencing buffer (provided with RNase T1) for 1 min at 95°C. Subsequently, 1 μl RNase T1 (0.1 u/μl) was added and incubation continued at 37°C for 5 min. OH ladders were generated by 5 min incubation of 0.2 pmol labelled RNA in alkaline hydrolysis buffer (provided with RNase T1) at 95°C. Reactions were stopped with 12 μl loading

buffer. Samples were denatured for 3 min at 95°C prior to separation on 6% polyacrylamide/7M urea sequencing gels in 1X TBE. Gels were dried and analysed using a PhosphorImager (FLA-3000 Series) and AIDA software (Raytest, Germany).

8.3.9. 30S ribosome toeprints

Toeprinting reactions were carried out as described (Hartz *et al.*, 1988; Udekwu *et al.*, 2005) with a few modifications. 0.2 pmol of an unlabelled *dppA* mRNA fragment (236 nt, T7 template amplified with JVO-1034/-1035), and 0.6 pmol of 5'end labelled primer JVO-1035 complementary to the *dppA* coding region were annealed. For inhibition analysis, 0.2, 0.6, 1 and 2 pmol of GcvB RNA or 1 pmol control RNA (MicA) or GcvB mutant RNAs were added. Nucleic acids were denatured in annealing buffer (10 mM Tris-acetate pH 7.6, 1 mM DTT, 100 mM potassium acetate) for 1 min at 95 °C and chilled on ice for 5 min, upon which Mg^{2+} acetate and all dNTPs were added to final concentrations of 10 mM and 0.5 mM, respectively. All subsequent incubation steps were at 37 °C. After 5 min incubation, 2 pmol of 30S ribosomal subunit (provided by Knud Nierhaus, Max Planck Institute for Molecular Genetics, Berlin, Germany; pre-activated for 20 min prior to the assay) were added. Following incubation for 5 min, uncharged $tRNA^{fMet}$ (10 pmol) was added, and incubations continued for 15 min. Reverse transcription was carried out by addition of 100 units of Superscript II and incubation for 20 min. cDNA synthesis was terminated with 100 μl stop buffer. Following phenol-chloroform extraction, alkaline hydolysis of template RNA at 90°C, and ethanol precipitation, cDNA was dissolved in 10 μl of loading buffer II (Ambion). Toeprint analysis on *gltI* 5' RNA (161 nt, T7 template amplified with primers JVO-1039/-1040) was performed using 5' end-labelled primer JVO-1775. See figure legend for final concentrations of other components. Sequencing ladders were generated with CycleReader^{TM} DNA Sequencing Kit according to the manufacturer's protocol on the same DNA template as used for T7 transcription and the same 5'-end-labelled primer as in the toeprinting reactions. cDNAs and sequence ladders were separated on a 6% polyacrylamide/ 7M urea gel. Autoradiograms of dried gels were obtained as above.

Stop buffer:
50 mM Tris-HCl pH 7.5
0.1% SDS
10 mM EDTA

8.3.10. *In vitro* translation assays

Translation reactions were carried out using Puresystem according to the manufacturer's instructions. 10 μl (Figs. 3.16B and D) or 20 μl (Fig. 3.17C) reactions contained, in addition to 70S ribosomes, mRNA template, Hfq and - where applicable - GcvB RNAs (see figure legends for final concentrations). Hfq dilutions were prepared in 1$\times$ dilution buffer.

1$\times$ Dilution buffer:
1$\times$ structure buffer
1% glycerol
0.1% Triton-X-100

Tables 8.9 and 8.10 list the details of DNA fragments used for T7 transcription. Before addition of puresystem mix, RNA was denatured for 1 min at 90°C and chilled on ice for 5 min. Hfq was mixed with mRNA (and sRNA), and pre-incubated for 10 min at 37°C. Puresystem mix was added and incubation continued at 37°C for the time given in the figure legend. Reactions were stopped with 4 vol of ice-cold acetone, kept on ice for 15 min, and proteins collected by centrifugation (10,000 g, 10 min, 4°C). Proteins were quantified by Western blot analysis with a monoclonal GFP antibody as previously published in Urban & Vogel (2007) and described in Section 8.2.4.5.

8.3.11. Quantitative RT-PCR

10 ml of a *Salmonella* culture grown to an OD_{600} of 0.4 (4 OD_{600} in total) were harvested, treated with 1/5 volume of stop solution (95% EtOH; 5% water-saturated phenol), snap-frozen in liquid nitrogen, and stored at -80°C until RNA extraction. RNA was isolated using the Promega SV total RNA purification kit as described in Section 8.2.5.2. RNA concentrations were determined on a Nanodrop machine. The relative amount of target mRNAs was determined by q-PCR using QuantitectTM SYBR Green RT-PCR Kit following manufacturer's instructions (Qiagen) in a 7900HT-RealTime-PCR System. For each reaction (25 μL final volume) 1 μl of RNA sample (100 ng / reaction) were mixed with 0.25 μl of primer pairs (0.5 μM final) and 12.5 μl of SYBR Green mix. For coupled cDNA synthesis and target gene amplification 0.25 μl of Quantitect RT mix was added. Each sample was assayed in triplicate for each run. Control RNA from wild-type cells was used to construct a standard curve for all inspected genes. Reaction conditions were: 30 min 50°C, 15 min 95°C, and 45 cycles at 94°C for 20 s, 60°C for 40 s, and 72°C for 40 s. Specific primer pairs for *Salmonella dppA* (JVO-1254/1255), *oppA* (JVO-1256/1257), *gltI* (JVO-1381/1382), *livJ* (JVO-1628/1629), *livK* (JVO-2326/2327), *argT* (JVO-2328/2329), and STM4351 (JVO-2330/2331) were designed using the PRIMER EXPRESS^{TM} software (Applied Biosystems). The *rpoA* gene (JVO-1340/1341) was used as an internal standard.

8.4. Methods: GcvB RNA, a global regulator of genes involved in amino acid metabolism

8.4.1. Strains, plasmids, and oligonucleotides

The *Salmonella enterica* serovar Typhimurium and *Escherichia coli* strains used in this section are listed in Table 10.1 in the Appendix. The complete list of DNA oligonucleotides used for cloning and as probes in hybridization is provided in Table 10.4 in the Appendix.

8.4.2. Plasmids

Plasmids that were used or constructed in this section are listed in Table 10.2 in the Appendix. Details of their construction, and insert sequences are given in Tables 8.11-8.14.

Table 8.11: Construction of additional GcvB RNA mutant plasmids.

Plasmid	Name	Template	Oligo 1	Oligo 2	Deletion of position
pKP30-1	pBAD-GcvB$_{\Delta R2}$	pKP1-1	JVO-0895	JVO-0896	136 - 144
pJL36-5	pP_L*gcvB*$_{\Delta R1\ \&\ \Delta R2}$	pJL22	JVO-0895	JVO-0896	66 - 89, 136 - 144
pJL57-1	pP_L*gcvB*$_{5'\Delta\ \&\ \Delta R2}$	pJL29-4	JVO-0895	JVO-0896	1 - 91, 136 - 144
pJL65-3	pP_L*gcvB*$_{\Delta SL2}$	pTP09	JVO-2856	JVO-2857	92 - 113
pJL66-12	pP_L*gcvB*$_{\Delta SL2\ \&\ \Delta SL3}$	pTP09	JVO-2856	JVO-2858	92 - 134
pFS127-2	pP_L*gcvB*$_{SL1,\ SL4\ \&\ SL5}$	pTP09	JVO-0745	JVO-0895	66 - 144
pFS129-2	pP_L*gcvB*$_{SL4\ \&\ SL5}$	pTP09	PLlacO-D	JVO-0895	1 - 144
pFS130-1	pP_L*gcvB*$_{R2,\ SL4\ \&\ SL5}$	pTP09	PLlacO-D	JVO-2989	1 - 134
pFS131-1	p*gcvB*$_{\Delta R1\ \&\ 3'\Delta T}$	pJL17-6	JVO-0745	JVO-0746	66 - 89, 135 - 206, and region downstream of *gcvB* terminator
pJL73-14	pP_L*gcvB*$_{SL1\ \&\ SL5}$	pTP09	JVO-0745	JVO-2986	66 - 177
pJL79-16	pP_L*gcvB*$_{M2,\ R2,\ SL4\ \&\ SL5}$	pFS130-1	JVO-3327	JVO-3328	1- 134, $G_{143}\rightarrow C$, $C_{158}\rightarrow G$
pJL78-11	pP_L*gcvB*$_{5'\Delta 12nt,\ SL1\ \&\ SL5}$	pJL73-14	PLlacO-D	JVO-2990	1 - 12, 66 - 177
pJL77-3	p*gcvB*$_{5'\Delta 12nt,\ \Delta R1\ \&\ 3'\Delta T}$	pFS131-1	JVO-0743	JVO-2990	1 - 12, 66 - 89, 135 - 206, and region downstream of *gcvB* terminator
pJL85-4	p*gcvB*$_{\Delta PL}$	pTP09	JVO-3355	JVO-1396	(-1) - (-35)
pSP9-1	pP_L*gcvB*$_{SL1\ \&\ SL5,\ C3}$	pJL73-14	JVO-3466	JVO-3467	$T_3\rightarrow C$, 66 - 177
pSP11-1	pP_L*gcvB*$_{SL1\ \&\ SL5,\ C8}$	pJL73-14	JVO-3464	JVO-3465	$G_8\rightarrow C$, 66 - 177
pSP10-1	pP_L*gcvB*$_{SL1\ \&\ SL5,\ G11}$	pJL73-14	JVO-3468	JVO-3469	$C_{11}\rightarrow G$, 66 - 177

8.4.2.1. sRNA plasmid construction

Plasmid pBAD-GcvB (pKP1-1) was constructed similarly as described for pBAD-RybB in Papenfort *et al.* (2006) but using primers JVO-0897 and pZE-XbaI for insert amplification on pTP05 which carries the *Salmonella gcvB* locus (292 bp upstream of the +1 site to 116 bp downstream of the terminator). Similarly, plasmid pJL3-15 and primers JVO-0897 and pZE-XbaI were used for amplification of the *gcvB*$_{\Delta R1}$ allele which was cloned under the control of the pBAD promoter analogous to pKP1-1 resulting in plasmid pKP2-6 (pBAD-GcvB$_{\Delta R1}$). Plasmid pKP30-1 (pBAD-GcvB$_{\Delta R2}$) was constructed by PCR amplification from plasmid pKP1-1 using Phusion Polymerase and primer pair JVO-0895/JVO-0896, DpnI digestion of template plasmid, and self-ligation of purified PCR product.

GcvB mutant plasmids were constructed via PCR amplification from the original plasmids using Phusion Polymerase, DpnI digestion of template plasmid, and self-ligation of purified PCR products. Primers and templates used for each mutant plasmid are listed in Table 8.11. Sequences of the mutant inserts are given in Table 8.12.

Table 8.12: Inserts of new GcvB mutant plasmids.

Black letters indicate the *gcvB* wild-type sequence, deleted parts of the individual GcvB mutants are marked in red, and the modified terminator/inserted nucleotides for the 3' deletion mutants are shown in green. Single nucleotide exchanges are indicated in blue uppercase letters.

Name	**Plasmid**	**Insert from +1 to end of *gcvB* terminator**	**Positions deleted as compared to wild-type *gcvB***
pP_L*gcvB*	pTP09	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	none
pP_L*gcvB*$_{\Delta R1}$	pJL22	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggtgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	66 - 89
pP_L*gcvB*$_{\Delta R2}$	pJL23	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	136 - 144
pP_L*gcvB*$_{5'\Delta}$	pJL29-4	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	1 - 91
p*gcvB*$_{3'\Delta T}$	pJL17-6	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcggaaacgctaccgttt ttttc	135 - 206; modification in terminator starts at position 121
pP_L*gcvB*$_{\Delta R1\ \&\ \Delta R2}$	pJL36-5	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggtgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	66 - 89, 136 - 144
pP_L*gcvB*$_{5'\Delta\ \&\ \Delta R2}$	pJL57-1	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	1 - 91, 136 - 144
pP_L*gcvB*$_{\Delta SL2}$	pJL65-3	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	92 - 113
pP_L*gcvB*$_{\Delta SL2\ \&\ \Delta SL3}$	pJL66-12	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	92 - 134

continued on next page

Name	**Plasmid**	**Insert from +1 to end of *gcvB* terminator**	**Positions deleted as compared to wild-type *gcvB***
$pP_L gcvB_{SL1, SL4 \& SL5}$	pFS127-2	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	66 - 44
$pP_L gcvB_{SL4 \& SL5}$	pFS129-2	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	1 - 144
$pP_L gcvB_{R2, SL4 \& SL5}$	pFS130-1	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	1 - 134
$pgcvB_{\Delta R1 \& 3'\Delta T}$	pFS131-1	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcggaaacgctaccgttt ttttttc	66 - 89, 135 - 206; modification in terminator starts at position 121
$pP_L gcvB_{SL1 \& SL5}$	pJL73-14	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	66 - 177
$pP_L gcvB_{M2, R2, SL4 \& SL5}$	pJL79-16	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctCtacatttaccctgtGtgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	1 - 134, $G_{143} \rightarrow C$, $C_{158} \rightarrow G$
$pP_L gcvB_{5' \Delta 12nt, SL1 \& SL5}$	pJL78-11	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	1 - 12, 66 - 177
$pgcvB_{5' \Delta 12nt, \Delta R1 \& 3'\Delta T}$	pJL77-3	acttcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcggaaacgctaccgttt ttttttc	1 - 12, 66 - 89, 135 - 206; modification in terminator starts at position 121
$pP_L gcvB_{SL1 \& SL5, C3}$	pSP9-1	acCtcctgagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	$T_3 \rightarrow C$, 66 - 177
$pP_L gcvB_{SL1 \& SL5, C8}$	pSP11-1	acttcctCagccggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	$G_8 \rightarrow C$, 66 - 177

continued on next page

Name	**Plasmid**	**Insert from +1 to end of *gcvB* terminator**	**Positions deleted as compared to wild-type *gcvB***
$pP_L gcvB_{SL1\&SL5,G11}$	pSP10-1	acttcctgagGcggaacgaaaagttttatcggaatgcgtgttctg atgggcttttggcttacggttgtgatgttgtgttgttgtgtttgc aattggtctgcgattcagaccacggtagcgagactacccttttc acttcctgtacatttaccctgtctgtccatagtgattaatgtagc accgccatattgcggtgctttttttt	C_{11}→G, 66 - 177

8.4.2.2. Construction of GFP-reporter plasmids

Translational GFP fusions to GcvB target mRNAs were constructed as described in Urban & Vogel (2007). For *cycA*, the promoter was mapped by 5' RACE as described in Section 8.2.5.7; the BseRI/NheI-digested cDNA product was cloned into the BsgI/NheI-digested fusion plasmid pXG-20. All other fusions were cloned into vector pXG-10. The details for GFP plasmid construction are listed in Table 8.13, and the inserts of all GFP fusions are listed in Table 8.14.

A shortened *cycA::gfp* fusion to the 10th amino acid was constructed by PCR amplification from the original plasmid pJL30-14 using Phusion polymerase and oligos JVO-3330 and JVO-0323, DpnI digestion of template plasmid, NheI digestion, and self-ligation of purified PCR products. This resulted in plasmid pJL83-2.

8.4.2.3. Transcriptomic experiments

RNA extraction and data generation were carried out with SALSA microarrays as described in Papenfort *et al.* (2006). RNA was isolated using the Promega SV total RNA purification kit as described in Section 8.2.5.2. The microarrays used in this study include PCR products of all the genes present in the sequenced *S. typhimurium* strain LT2. In addition, 229 genes specific to *S. typhimurium* strain SL1344 were added. Details of all the amplicons can be found at http://www.ifr.ac.uk/Safety/MolMicro/pubs.html. The experimental design involves the use of *Salmonella enterica* serovar Typhimurium genomic DNA as the co-hybridized control for one channel on all microarrays. This method has the advantage of allowing the direct comparison of multiple samples. Total RNA and chromosomal DNA were labelled by random priming according to the protocols described at the IFR (Institute of Food Research, Norwich) website[4]. Briefly, 16 μg RNA were reverse transcribed and labelled with Cy3-conjugated dCTP (Pharmacia) using 200 units of StrataScript and random octamers (Invitrogen). Chromosomal DNA (400 ng) was labelled with Cy5-dCTP using the Klenow fragment. After labelling, each Cy3-labelled cDNA sample was combined with Cy5-labelled chromosomal DNA and hybridized to a microarray overnight at 65ºC. After hybridization, slides were washed and scanned using a GenePix 4000A scanner. Fluorescent spots and the local background intensities were identified and quantified using Bluefuse software (BlueGnome, Oxford). To compensate for unequal dye incorporation, data centring to zero was performed for each block (one block being defined as the group of spots printed by the same pin). Microarray data were analysed using GeneSpring 7.3 (Agilent) and genes were considered to be differentially expressed if they displayed $\geq$2-fold changes in all replicates and were statistically significantly different.

[4] www.ifr.ac.uk/safety/microarrays/protocols.html

Table 8.13: Details of new GFP-fusion plasmids construction.

Target gene	Plasmid trivial name	Plasmid original name	Backbone	Oligos	Insert digested with	5' RACE end/TAP specific	Ecocyc promoter[a]	Insert 5' end [bp]	Insert 3' end [bp]	Total insert size [bp]	GFP fused to aa	Signal sequence [aa][b]	Fluorescence on plate
cycA	*cycA::gfp*	pJL30-14	pXG-20	JVO-0367 × JVO-1275	BseRI/NheI	Yes/ -	-	-79	+57	136	19	maybe Tat	+++
ydgR (tppB)	*ydgR::gfp*	pJL70-9	pXG-10	JVO-2850 × JVO-2851	BfrBI/NheI	-	-98	-96	+45	141	15	no	+
ygjU (sstT)[c]	*ygjU::gfp*	pFS133-3	pXG-10	JVO-2971 × JVO-3087	BfrBI/NheI	-	-	-81	+21	102	7	no	+
yaeC	*yaeC::gfp*	pFM27-1	pXG-30	JVO-2058 × JVO-2059	BfrBI/NheI	-	operon *metNIQ*	-41	+36	77	12	21	++
gdhA	*gdhA:gfp*	pJL69-5	pXG-10	JVO-2806 × JVO-2807	BfrBI/NheI	-	-63	-64	+33	97	11	no	+
asd[c]	*asd::gfp*	pFS116-1	pXG-10	JVO-2969 × JVO-2970	BfrBI/NheI	-	-	-101	+27	128	9	no	weak
lrp	*lrp::gfp*	pFS103-3	pXG-10	JVO-2800 × JVO-2801	BfrBI/NheI	-	-267	-225	+45	270	15	no	+
ilvC	*ilvC::gfp*	pJL68-1	pXG-10	JVO-2804 × JVO-2805	BfrBI/NheI	-	-58	-68	+33	101	11	no	+
iciA (argP)	*iciA::gfp*	pFS121-1	pXG-10	JVO-2973 × JVO-2874	BfrBI/NheI	-	-23	-27	+36	63	12	no	+
brnQ[d]	*brnQ::gfp*	pFS105-3	pXG-10	JVO-2842 × JVO-2843	BfrBI/NheI	-	-	-174	+30	204	10	26	+
ilvE	*ilvE::gfp*	pSP25-7	pXG-10	JVO-3388 × JVO-3389	BfrBI/NheI	-	-96	-94	+36	130	12	no	++
thrL	*thrL::gfp*	pSP20-1	pXG-10	JVO-3378 × JVO-3379	BfrBI/NheI	-	-42	-42	+57	99	19	no	++
ybdH	*ybdH::gfp*	pSP21-2	pXG-10	JVO-3380 × JVO-3381	BfrBI/NheI	-	-	-71	+30	101	10	no	weak
ndk	*ndk::gfp*	pFS115-2	pXG-10	JVO-2965 × JVO-2808	BfrBI/NheI	-34	-	-35	+39	74	13	no	+
serA	*serA::gfp*	pFS117-1	pXG-10	JVO-2967 × JVO-2968	BfrBI/NheI	-	-139/ -46	-51	+30	81	10	no	+

[a] Promoter in *E. coli* based on EcoCyc `www.ecocyc.org` annotation.

[b] Signal peptide predicted with SignalP 3.0 server (`http://www.cbs.dtu.dk/services/SignalP/`); (Bendtsen *et al.*, 2004).

[c] Promoter taken from Hfq-CoIP 454 data (Sittka *et al.*, 2008).

[d] Promoter taken from Ohnishi *et al.* (1988).

Table 8.14: Inserts of new GFP-fusion plasmids. *Salmonella* gene sequences are capitalized, in which black letters correspond to 5' UTR parts and red letters to ORF parts. BfrBI sites of pXG-10 and NheI sites that were used for cloning are highlighted in bold in magenta and green, respectively.

GFP fusion	Insert
cycA::gfp	GTCCTGATAACAGGATCGTCGTATCATAGACCAAAGGCCGTAGAGCCCGCACAACACAGA CAGGTACAGGAAGAAAAACATGGTAGATCAGGTAAAAGTCGCAGCCGACGAACAGGCTCC GGCTGAACAGTCGCTA**gctagc**
cycA_10th::gfp	GTCCTGATAACAGGATCGTCGTATCATAGACCAAAGGCCGTAGAGCCCGCACAACACAGA CAGGTACAGGAAGAAAAACATGGTAGATCAGGTAAAAGTCGCAGCCGAC**gctagc**
ydgR::gfp	**atgcat**GCCGTTTCCCCTCCAATATAACAATCGGACGGATGAGTCTGACTCATCACGCGC CAGACAATCCCTGTTAATACGGGGCGTAAAAAAGAGGTAAAAGTGTCTACTGCAAACAAA AAACCAACTGAAAGCGTCAGTTTGAAC**gctagc**
ygjU::gfp	**atgcat**GCAAACACTTTGTTACATCCTGAAAGATGCGCCGTAAGAGCGTGCAGGGGATGA CCAGCAACACAATACAAGGAATATGAAATGGCTACGCAACGAGCATCA**gctagc**
yaeC::gfp	**atgcat**TAACGTTAAACACAACACAAATACTCATTAAGGAAATAACGATGGCGTTCAAAT TCAAAACCTTTGCGGCAGTGGGT**gctagc**
gdhA::gfp	**atgcat**GCAAATACATATTCTGATAAAACGCAAATACAACCACATTAATATATAAGAGGT TTTTATATCTATGGATCAGACATGTTCTCTGGAATCGTTCCTC**gctagc**
asd::gfp	**atgcat**TTAATTTCACTTGCGACTTTGGCTGCTTTTTGTATGGTGAAGGATGCGCCACAG ATACTGGCGCGCATACACAGCACATCTCTTTGCAGGAAAAAAACGCTATGAAAAATGTTG GTTTTATCGGCTGG**gctagc**
lrp::gfp	**atgcat**GTAAATACCATGTTTACCGGGCTAGTGAAATCTACGCATGGCGTGGACAGACGC CATTCGTGATGTCGATAGCTGCCGCGAGGCAACGGTCTTCTCACCATAGACCAGGCATTG CGCGCCGTTAATCCCTCTGGGTTTCGGTCTATCGTGATGGGCAGCGACTCTGAACAGTGA TGTGAGTAGAGTCAGGCAGGAGTAGGGAAGGAATACAGAGAGACAATAATAATGGTAGAT AGCAAGAAGCGCCCTGGCAAAGATCTCGACCGTATC**gctagc**
ilvC::gfp	**atgcat**ATTCGCACAGATAGCAATCTGTAAACCGAACAATAAGCGCGACACACAACATCA CGGAGTACACCATCATGGCTAACTACTTTAATACACTGAATCTGCGC**gctagc**
iciA::gfp	**atgcat**AAAAAATAACAGGAGCATGACACAACAATGAAACGTCCGGACTACAGAACACTA CAGGCGCTG**gctagc**
brnQ::gfp	**atgcat**TCAGGTGCTGTCATTACGACTGCATTTGCGCGGTAAATCGAAAAACAATTCTTC GCCGCATCGGTCCGGGAGCTTTTCCCGCTGAAATTGATAAAAAACGCCGCTTTATAATCC TCCGGGAAAGGCAAAAATTTTAATGTTTACAACATCACCATCCACAGGCAGTAAGACTTT ATGACCCATCAGTTAAAATCGCGCGATATC**gctagc**
ilvE::gfp	**atgcat**AGTTAAGTAAACTGGTAGATGTTGCGCATGTCGCGATCTGCCAGAGCGCTGCCA CATCACAACAAATCCGCGCCTGAGCGCAAAAGGAAGAAAAATGACGACGAAAAAAGCTGA TTATATTTGGTTCAAT**gctagc**
thrL::gfp	**atgcat**ATACAAGACAGACAAATAAAAATGACAGAGTACACAACATCCATGAACCGCATC AGCACCACCACCATTACCACCATCACCATTACCACAGGTAACGGT**gctagc**
ybdH::gfp	**atgcat**ATTTGGCAATCAAGACGTTTAGATGTCTAAATATAATAACAACGGTGAGAAGAC CCTAAGGACAACACAACATGAACCACACTGAGATCCGCGTCGTTACC**gctagc**
ndk::gfp	**atgcat**CTGACATAACAACAGAACATATTTCAGAGGTAAACATGGCTATTGAACGGACTT TTTCCATCATTAAACCCAAC**gctagc**
serA::gfp	**atgcat**GATGCAAATCCACACACAACATCAGATGGCAAAAAAGACAGGATCGGGGAAATG GCAAAGGTATCGCTGGAGAAAGATAAA**gctagc**

8.4.3. RNA and protein detection

RNA preparation and Northern analysis as well as GFP fusion and GroEL protein detection were performed as described in Sections 8.2.5 and 8.3.4. Colony fluorescence imaging was carried out as described in Section 8.3.5.

For FACS analysis, *E. coli* strains carrying *gfp* fusion plasmids were grown to stationary phase in 3 ml liquid cultures (inoculated from single colonies) in LB medium (supplemented with the appropriate antibiotics). After 14 h, bacteria from 1 ml culture volume were harvested by centrifugation for 2 min at 8,000 rpm and 4°C. After removal of supernatants, bacteria were resuspended in 500μl 2% PFA (paraformaldehyde) in 1×PBS buffer (pH 7.4) for fixation and stored for up to five days in the dark at 4°C until FACS analysis. Prior to FACS measurements, samples were diluted 1:250 in 1×PBS buffer.

10× PBS (phosphate-buffered saline) buffer, pH 7.4:
137 mM NaCl
2.7 mM KCl
4.3 mM Na_2HPO_4*2 H_20
1.4 mM KH_2PO_4

To determine reporter activities of single cells, a BD FACSCantoTM Flow Cytometer equipped with a blue excitation source (air-cooled, 20 mW solid state 488 nm laser) was used to measure forward angle light scatter (FSC), side scatter (SSC), and the fluorescence of the cells (FITC). The instrument settings were in logarithmic mode: FSC-H: 516, SSC-A: 626; FITC-A (GFP): 753. GFP fluorescence intensity was measured for 30,000 events (maximum threshold of 10,000 events/sec). All FACS analyses were done in duplicates. Data analysis was carried out using FCS Express software, version 3 (De Novo Software).

8.4.4. *In vitro* structure probing and 30S ribosome toeprinting

DNA templates carrying a T7 promoter sequence for *in vitro* transcription were generated by PCR. Primers and sequences of the T7 transcripts are included in Tables 8.9 and 8.10 in Section 8.3.6. RNA was *in vitro* transcribed, quality-checked and labelled at the 5' end as described in Section 8.2.5.8. Secondary structure probing and mapping of RNA complexes was conducted on 5'-end-labelled RNA as described in Section 8.3.8.

Toeprinting reactions were carried out as described in Section 8.3.9. Specifically, 0.2 pmol of an unlabelled *cycA_10th::gfp* mRNA fragment (171 nt, T7 template amplified with JVO-1274/-1976) and 0.5 pmol of 5'end labelled primer JVO-1976 complementary to the *gfp* coding region were annealed. For inhibition analysis, 0.2, 1 and 2 pmol of GcvB RNA or 2 pmol control RNA (MicA) or GcvB mutant RNAs were added.

8.4.5. Motif detection using `MEME` and `MAST`

The *gfp* fusion sequences of the seven old targets (*dppA*, *oppA*, *livJ*, *livK*, *argT*, STM4351, and *gltI*, see Table 8.7) and seven new targets from the microarray analysis (*ydgR*, *ygjU*, *yaeC*, *gdhA*, *asd*,

lrp, and *cycA*, see Table 8.14) were used as input sequences for MEME[5] motif identification (Bailey *et al.*, 2006). The following parameters were defined: number of different motifs: 10; minimum number of sites: 8; maximum number of sites: 14; minimum motif width: 6; maximum motif width: 25.

The position-specific weight matrix for the C/A-rich GcvB target derived from the MEME search was then used as input for MAST[6] (Bailey & Gribskov, 1998) searches in a database composed of the 5' regions (-70 to +30 according to the start codon) of all annotated *Salmonella* ORFs. These were extracted as a multi-Fasta file from the *Salmonella typhimurium* LT2 genome sequence and annotation (NC_003197) using own unpublished Perl[7] scripts.

8.4.6. Prediction of sRNA-target mRNA duplexes

GcvB-target mRNA complexes were predicted with RNAhybrid (Rehmsmeier *et al.*, 2004). 5' regions (-70/+30 nt of annotated start codon) of 246 *Salmonella* LT2 genes, where the C/A motif was predicted by MAST searches, were used as target sequence input. An extended GcvB region R1 (TTGGCTTACGGTTGTGATGTTGTGTTGTTGTGTTTGCAATTGGTCTGCG) was used as miRNA input. Two sets of interactions were predicted: one without limitations for bulge/internal loops and another where these loops were restricted to a length of one nucleotide. Interactions that have no helix with at least nine subsequent base-pairs were sorted out manually from the prediction lists. Furthermore, target interactions were predicted for the GcvB R1 sequence with the -70/+30 regions of all *Salmonella* mRNAs.

8.5. Methods: Hfq-coIP in *Salmonella*

8.5.1. Bacterial strains, growth, and oligodeoxynucleotides

The *Salmonella enterica* serovar Typhimurium strains used in this section are listed in Table 10.1 in the Appendix. The complete list of DNA oligonucleotides used as hybridization probes is provided in Table 10.9 in the Appendix. For early stationary phase (ESP) cultures, 30 ml L-broth in a 100 ml flask was inoculated 1/100 from overnight cultures and incubated at 37°C, 220 rpm to an optical density of 2.

8.5.2. RNA isolation and Northern blot analysis

RNA was prepared by hot phenol extraction as described in Mattatall & Sanderson (1996) and Section 8.2.5.3, followed by DNase I treatment. After separation on 5% polyacrylamide (PAA) gels containing 8.3 M Urea, RNA was transferred onto Hybond-XL membrane. 5 or 10 μg RNA was loaded per sample. For detection of new transcripts γ^{32}P-ATP end-labelled oligodeoxyribonucleotides were used (see Table 10.9 in the Appendix).

[5] meme.sdsc.edu/meme4_1/cgi-bin/meme.cgi

[6] meme.sdsc.edu/meme4_1/cgi-bin/mast.cgi

[7] www.perl.org

8.5.3. Hfq co-immunoprecipitation, cDNA synthesis and high-throughput pyrosequencing (HTPS)

Strains SL1344 and JVS-1338 (*hfq*^{FLAG}) were grown in L-broth under normal aeration at 37°C to early stationary phase. Co-immunoprecipitation was carried out using the protocol published in Pfeiffer *et al.* (2007). For pyrosequencing, samples of two independent pull down experiments were used. cDNA cloning and pyrosequencing was performed as described for the identification of eukaryotic microRNA (Berezikov *et al.*, 2006) but omitting size-fractionation of RNA prior to cDNA synthesis. The cDNA libraries were constructed by *vertis Biotechnology AG*[8] and sequenced at the Max Planck Institute for Molecular Genetics in Berlin, Germany.

8.5.4. Hardware and software used for 454 sequencing analysis

This following programming languages, programs, tools, and existing methods were used for the analysis of 454 sequencing data:

8.5.4.1. Development environment and programming language

The operating system Fedora[9] GNU/Linux core 8 with kernel 2.6.x was used during the analysis of 454 sequencing data. Programs were run on a computer with an Intel(R) Pentium(R) M processor 2.0 GHz and 2 GB RAM.

Scripts for analysis of 454 data were programmed in Perl[10](Practical Extraction and Report Language). Perl is both a very simple and a high-level language suited especially for string handling, pattern recognition in data and texts, and rapid prototyping. It allows easy programming solutions for data evaluations with far less programming effort than in C or Java because, *e. g.*, Perl automatically manages the memory allocation. Perl v.5.8.8 was used for the implementation of all analysis programs of this thesis because of its above mentioned strengths in the evaluation of data files.

8.5.4.2. Tools

BLAST[11] (Basic Local Alignment Search Tool) provides a method for rapid searching in nucleotide and protein databases (Altschul *et al.*, 1990). Here, `blastn` (for standard nucleotide-nucleotide BLAST) of the `WU-BLAST`[12] package was used for mapping of 454 reads. The BLAST algorithm uses an indexed table or dictionary of short subsequences called words for both the query sequence and the database to find similar sequences. It localizes rapidly initial exact matches to the query words by simply looking up a particular word in the database dictionary. With these initial matches serving as starting points for longer alignments a final gapped and scored alignment is generated in several steps.

[8] `http://www.vertis-biotech.com/`
[9] `fedoraproject.org`
[10] `http://www.perl.com`
[11] `http://www.ncbi.nlm.nih.gov/BLAST/`
[12] `http://blast.wustl.edu/`

Patscan[13] is a pattern matcher which searches protein or nucleotide sequence archives for instances of a pattern which is provided as input (Dsouza *et al.*, 1997). Patscan was used for 5' end linker and polyA tail identification during the clipping step.

The **Integrated Genome Browser**[14] (IGB, pronounced ig-bee) is an application intended for visualization and exploration of genomes and corresponding annotations from multiple data sources. The IGB is also a part of the open source GenoViz[15] project. Links to all documentation and Java source code for IGB, including the most recent version of this document, can be found on the homepage of the GenoViz project. The IGB can work with three distinct types of data: annotations, graphs, and genomic sequences. Annotations indicate the known or suspected locations of features such as mRNAs, rRNAs, promoter regions, pseudogenes, and so forth. Annotation data can be directly loaded from files. The IGB was chosen for visualization of mapped 454 reads, as it is able to display graphs. Graphs indicate scores or other numeric values as a function of genomic position and can be displayed as some form of plot (x,y-plot, bar plot, etc.).

8.5.5. Analysis and visualization of pyrosequencing results

A detailed description of the developed analysis tools, the analysis workflow and visualization of pyrosequencing results is given in Section 5.1.2.

In brief, after 5'end linker and polyA-tail clipping from the initial pyrosequencing results, all inserts $\geq$18 nt of the Hfq-coIP and coIP-Ctr libraries were separately mapped to the *Salmonella* LT2 genome (NC_003197.fna) using `WU-BLAST`. From the resulting `BLAST` positions one graph for each strand of the *Salmonella* chromosome was calculated, where the number of cDNA hits for each nucleotide position was plotted. To compare the graphs of the Hfq-coIP and coIP-Ctr, the graphs were normalized to number of mapped reads. Following upload of the *Salmonella* genome sequence and annotation (NC_003197.fna and NC_003197.gff), the two graphs for each library were loaded into the Integrated Genome Browser (IGB) of Affymetrix (version 4.56).

8.6. Methods: Deep sequencing reveals the primary transcriptome of *Helicobacter pylori*

8.6.1. Bacterial strains and oligodeoxynucleotides

Throughout the whole study, *Helicobacter pylori* strain 26695 was used. *H. pylori* strain 26695 (CIP 106780) was purchased from the Collection of the 'Institut Pasteur' (CIP).

The complete lists of DNA oligonucleotides used as hybridization probes or for 5' RACE analysis are provided in Tables 10.14 and 10.15 in the Appendix.

[13] `http://www-unix.mcs.anl.gov/compbio/PatScan/HTML/patscan.html`

[14] `http://www.affymetrix.com/partners_programs/programs/developer/tools/igbsource_terms.affx`

[15] `http://genoviz.sourceforge.net/`

8.6.2. Biocomputational prediction of sRNAs

A bioinformatics-based approach similar to Chen *et al.* (2002) and Argaman *et al.* (2001) was taken for the prediction of novel sRNA candidates in *Helicobacter pylori* 26695 (see Section 2.3.3). First, all intergenic regions ≥ 60 nt were extracted from the genome sequence based on the annotations specified in the *H. pylori* 26695 Genbank file `NC000915.gb` that can be downloaded from `NCBI`[16]. Intergenic regions are defined as regions where no gene is annotated on either of the two strands. The resulting list of 636 intergenic regions ≥ 60 in *H. pylori* 26695 was then scanned for orphan promoter/ terminator pairs. Rho-independent transcription terminators were predicted like in Chen *et al.* (2002) using the `RNAMotif` (Macke *et al.*, 2001) descriptor file specified in Lesnik *et al.* (2001). From the resulting list of 109 predicted terminators with energy scores between -1.03 and -9.66, only terminators with a threshold value less than -3.0 were considered for further analysis. A less sringent threshold compared to Chen *et al.* (2002) was used, as it was unclear whether *Helicobacter* Rho-independent terminators are as stable as those from *E. coli*. Thus, the terminator predictions ended up with a list of 56 possible terminators.

From this list, all terminators that are close to flanking genes so that either this terminator belongs to the flanking gene or the space to the flanking gene is not long enough to harbour a promoter sequence were removed. This resulted in 29 predicted terminators. For promoter predictions, only regulatory elements were used that were described for the *H. pylori* σ^{80} (RpoD) family of promoter signals (Vanet *et al.*, 2000) which is related to σ^{70} from *E. coli*. Promoters were predicted by pattern searches with `PatScan` (Dsouza *et al.*, 1997). Only promoter/ terminator pairs on the same strand and correct orientation within 40 - 400 nt were considered as putative sRNA candidates. No appropriate promoter could be predicted for eleven terminators. After removal of six putative, non-annotated ORFs using `ORF finder`[17] and `blastx` from the list of 18 promoter/terminator pairs, a search for conservation of the remaining 12 sRNA candidates in the *H. pylori* strain J99 was conducted. This restrictive search led to a final list of six sRNA candidates (Table 10.10 in the Appendix) that were selected for experimental verification, whereof one is HP26695 specific. The alignments for sRNA candidates of *H. pylori* 26695 and J99 can be found in Figures 10.7 to 10.11 in the Appendix.

8.6.3. *Helicobacter* growth

Bacteria were grown on columbia agar plates supplemented with 7% laked horse blood and the Dent selective supplement (Oxoid, Basingstoke, Hampshire, England) for 24 to 48 hours at 37°C in anaerobic jars under microaerobic conditions (10% CO_2, 6% O_2) generated by CampyGen bags. For liquid cultures, plate-grown bacteria were harvested and resuspended to a final OD_{600} of 0.08 in a flask containing 25 mL of Brain Heart Infusion (BHI) medium, supplemented with 10% foetal calf serum (FCS) and Dent selective supplement. A starter culture was prepared by incubating the flasks at 37°C, under microaerobic atmosphere and agitation at 120 rpm to an OD_{600} of 1.5.

[16] `http://www.ncbi.nlm.nih.gov/`

[17] `http://www.ncbi.nlm.nih.gov/gorf/gorf.html`

8.6.3.1. Acid stress.

Cells from the starter culture were diluted to an OD_{600} of 0.08 in three flasks containing fresh BHI medium, supplemented with 10% foetal calf serum (FCS) and Dent's antibiotics, and then incubated for 15 h at 37°C in jars under microaerobic conditions. After 15 hours of growth, one flask was used to estimate the volume of HCl (3,7%) required to lower the pH to 5.5. Then, the estimated volume of HCl or sterilized H_20 was added to the acid stress flask and the control flask, respectively. Both flasks were incubated for 30 more minutes, at 37°C, under microaerobic conditions. After incubaction, bacterial growth was stopped by adding 3 mL stop solution (95% ethanol/5% phenol) to 25 mL of bacterial culture. Cells were harvested and centrifuged for 10 minutes at 3,000 rpm and 4°C. Then, the supernatant was removed and pellets were stored at -80°C.

8.6.3.2. Co-infection with human cells.

Huh7 and AGS cell lines were cultivated in flasks containing Dulbecco's Modified Eagle's Medium (DMEM) or Ham's medium, respectively, at 37°C, in a humidified atmosphere of 5% CO_2, 95% air. Before the co-infection experiment, the medium was removed from the flasks, cells were washed with PBS, and fresh Ham's medium was added in both flasks containing AGS and Huh7 adherent cells, and in one more flask containing no cells (plastic control). Then AGS, Huh7 and control flasks were inoculated with a *H. pylori* suspension from 48 h plate-grown cultures, at a multiplicity of infection (MOI) of 240, and incubated at 37°C for 7 h. To collect bacteria adherent to host cells, first cells were removed from plastic substrate by using a cell scraper. Then, stop solution was added to each flask, and the whole suspension (composed of supernatant bacteria and released cell-adherent bacteria) was harvested, centrifuged at 4°C, for 10 minutes at 3.000 rpm, and pellets were stored at -80°C.

8.6.4. RNA extraction, Northern blot analysis, and 5' RACE analysis

Frozen pellets from acid stress and co-infection experiments were thawed on ice and resuspended in lysis solution containing 800 μl of 0.5 mg/mL lysozyme in TE buffer (pH 8.0), and 80μl 10% SDS. Bacterial cells lysis was done by placing the samples in a water bath for 1-2 minutes at 64°C. Total RNA was then extracted from the lysates by using the hot-phenol method described in Blomberg *et al.* (1990).

For Northern Blot analysis, 3 to 20 μg RNA was loaded per sample. After separation on 6% polyacrylamide (PAA) gels containing 8.3 M urea, RNA was transferred onto Hybond-XL membranes, and membranes were hybridized with γ^{32}P-ATP end-labelled oligodeoxyribonucleotides probes listed in Table 10.14 in the Appendix.

5' RACE analysis of *Helicobacter* genes was done following previously published protocols (Argaman *et al.*, 2001; Urban & Vogel, 2007) and as described in Section 8.2.5.7. Gene specific antisense oligos that were used for 5' end mapping are listed in Table 10.15 in the Appendix.

8.6.5. Depletion of processed RNAs and construction of cDNA libraries

Total RNA from *Helicobacter* was first treated with DNase I. For depletion of processed transcripts, 7 μg of *Helicobacter* RNA were incubated for 60 min at 30°C with Terminator^{TM} 5'-Phosphate-

Dependent Exonuclease (TEX) or in buffer alone. Following P:C:I extraction, RNAs were precipitated overnight by addition of 2.5 volumes of an ethanol:sodium acetate (pH 5.2) mixture (30:1 v/v). Afterwards, RNAs were treated with TAP (tobacco acid pyrophosphatase) for one hour at 37°C to generate 5'-mono-phosphates for linker ligation, followed by an additional P:C:I extraction and ethanol precipitation step.

cDNA cloning and 454 pyrosequencing was performed as described for the identification of eukaryotic microRNA (Berezikov *et al.*, 2006) but omitting size-fractionation of RNA prior to cDNA synthesis. All cDNA libraries were constructed by *vertis Biotechnology AG*[18]. In brief, equal amounts of -/+ TEX treated RNA were poly(A)-tailed using poly(A) polymerase, followed by ligation of an RNA adapter to the 5'-phosphate of the small RNAs. First-strand cDNA synthesis was then performed using an oligo(dT)-adapter primer and M-MLV-RNaseH$^-$ reverse transcriptase. Incubation temperatures were 42°C for 20 min, ramp to 55°C followed by 55°C for 5 min. The resulting cDNAs were then PCR-amplified to 20-30 ng/μl using a high fidelity DNA polymerase. Each library contains a specific 4-mer barcode sequence which is attached to the 5'-end of the cDNAs.

Helicobacter cDNA libraries were prepared for -/+ TEX treated RNA samples from five growth conditions: culture in BHI media to mid-log phase (C-/+ libraries), and following 30 min acid stress (AS-/+); growth in cell culture flasks in the absence (PL-/+) or presence of two eukaryotic cell types, *i. e.* AGS human gastric epithelia cells (AGS-/+) as an infection model and Huh7 cells (Huh7-/+) as negative control.

8.6.6. 454 sequencing and read mapping

In total, 2.15 million reads were sequenced for the acid stress libaries and 1.79 million reads for the infection libraries. This resulted in 220,000-530,000 cDNAs per library (a total of $\approx$3.7 million cDNAs, see Table 8.15) on a Roche FLX sequencer. The cDNA libraries were sequenced at the Max-Planck-Institute for Molecular Genetics in Berlin, Germany, and in collaboration with Roche Diagnostics GmbH in Penzberg, Germany.

First 5'end linker sequences were clipped using the Perl script `clip_linker.pl`, which was developed for the analysis of the Hfq-coIP libraries described in Section 5.1.2. Afterwards, 5'-linker clipped reads were mapped to the *Helicobacter pylori* 26695 genome using the program `segemehl` which is based on an error-tolerant suffix array method (Hoffmann *et al.*, 2009, *submitted*). For this mapping method, clipping of tailing sequences is not necessary as they will be removed during the mapping step. However, for very short sequences the polyA tail often leads to mapping errors. Therefore, a filtering step was introduced which removed all sequences with an A-content of more than 70% (see Table 8.15). For these sequences, the polyA tail was clipped separately. Of the clipped reads, all sequences $\geq$12 nt were mapped again with `segemehl`. This procedure allowed mapping of also very short sequences of at least 12 nucleotides.

For each library, the mapped read length distribution (see Figure 6.9 in Section 6.1.3) was calculated for all mapped reads using a modified version of the Perl script `lengths_statistics.pl` described in Section 5.1.2. The modified script recognizes the tabular mapping output generated by `segemehl` instead of the tabular `BLAST` output generated by `WU-Blast`.

[18] `http://www.vertis-biotech.com/`

Table 8.15: Mapping statistics of ten *Heliobacter* cDNA libraries.

cDNA Library	Total number of 5'end clipped sequences	Sequences removed by dustbin filter (>70% 'A's)	< 12 nt in dustbin	≥12 nt in dustbin for remapping	No match	Total number of mapped reads	Percent mapped reads [%]	Mapped locations	Total number of uniquely mapped reads	Percent uniquely mapped reads [%]	Mapped nucleotides
C -	528373	25952	7515	18437	75887	444971	84,2	793725	114505	25.7	24387787
C +	528169	34403	14984	19419	90256	422929	80,1	618441	257645	60.9	27799949
AS -	427455	24055	8265	15790	63905	355285	83,1	518099	220089	61.9	26336292
AS +	540133	37263	18019	19244	108420	413694	76,6	546118	311267	75.2	27419459
PL -	268841	15781	5130	10651	61007	202704	75,4	319243	98682	48.7	14177628
PL +	315309	14680	6602	8078	57063	251644	79,8	377525	153904	61.2	18783183
AGS -	280713	22972	4547	18245	93029	183137	65,2	293503	85227	46.5	13337457
AGS +	223705	21794	5280	16214	78116	140309	62,7	204815	90533	64.5	9996795
Huh7 -	266621	9795	865	8930	66904	198852	74,6	326304	84105	42.3	27012137
Huh7 +	308759	12259	657	11602	100704	207398	67,2	326898	107017	51.6	27601375
SUM	3688078	218954	71864	146610	795291	2820923	76,5	4324671	1522974	54.0	216852062

CHAPTER 9

REFERENCES

Aarons, S., Abbas, A., Adams, C., Fenton, A. & O'Gara, F. (2000). A regulatory RNA (PrrB RNA) modulates expression of secondary metabolite genes in *Pseudomonas fluorescens* F113. *J. Bacteriol.*, 182(14), 3913–3919.

Abouhamad, W. N., Manson, M., Gibson, M. M. & Higgins, C. F. (1991). Peptide transport and chemotaxis in *Escherichia coli* and *Salmonella typhimurium*: characterization of the dipeptide permease (Dpp) and the dipeptide-binding protein. *Mol. Microbiol.*, 5(5), 1035–1047.

Afonyushkin, T., Vecerek, B., Moll, I., Bläsi, U. & Kaberdin, V. R (2005). Both RNase E and RNase III control the stability of *sodB* mRNA upon translational inhibition by the small regulatory RNA RyhB. *Nucleic Acids Res.*, 33(5), 1678–1689.

Aiba, H. (2007). Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.*, 10(2), 134–139.

Aiba, H., Adhya, S. & de Crombrugghe, B. (1981). Evidence for two functional *gal* promoters in intact Escherichia coli cells. *J. Biol. Chem.*, 256(22), 11905–11910.

Akada, J. K., Shirai, M., Takeuchi, H., Tsuda, M. & Nakazawa, T. (2000). Identification of the urease operon in *Helicobacter pylori* and its control by mRNA decay in response to pH. *Mol. Microbiol.*, 36(5), 1071–1084.

Alix, E. & Blanc-Potard, A. (2009). Hydrophobic peptides: novel regulators within bacterial membrane. *Mol. Microbiol.*, 72(1), 5–11.

Alkan, C., Karakoç, E., Nadeau, J. H., Sahinalp, S. C. & Zhang, K. (2006). RNA-RNA interaction prediction and antisense RNA target search. *J. Comput. Biol.*, 13(2), 267–282.

Alm, R. A., Ling, L. S., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C., deJonge, B. L., Carmel, G., Tummino, P. J., Caruso, A., Uria-Nickelsen, M., Mills, D. M., Ives, C., Gibson, R., Merberg, D., Mills, S. D., Jiang, Q., Taylor, D. E., Vovis, G. F. & Trust, T. J. (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, 397(6715), 176–180.

Altier, C., Suyemoto, M. & Lawhon, S. D. (2000). Regulation of *Salmonella enterica* Serovar Typhimurium invasion genes by *csrA*. *Infect. Immun.*, 68(12), 6790–6797.

Altier, C., Suyemoto, M., Ruiz, A. I., Burnham, K. D. & Maurer, R. (2000). Characterization of two novel regulatory genes affecting *Salmonella* invasion gene expression. *Mol. Microbiol.*, 35(3), 635–646.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Altuvia, S. (2007). Identification of bacterial small non-coding RNAs: experimental approaches. *Curr. Opin. Microbiol.*, 10(3), 257–261.

Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L. & Storz, G. (1997). A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell*, 90(1), 43–53.

Altuvia, S., Zhang, A., Argaman, L., Tiwari, A. & Storz, G. (1998). The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J.*, 17(20), 6069–6075.

Anantharaman, V., Koonin, E. V. & Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.*, 30(7), 1427–1464.

Ando, Y., Asari, S., Suzuma, S., Yamane, K. & Nakamura, K. (2002). Expression of a small RNA, BS203 RNA, from the *yocI-yocJ* intergenic region of *Bacillus subtilis* genome. *FEMS Microbiol. Lett.*, 207(1), 29–33.

André, G., Even, S., Putzer, H., P.Burguière, Croux, C., Danchin, A., Martin-Verstraete, I. & Soutourina, O. (2008). S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res.*, 36(18), 5955–5969.

Ansong, C., Yoon, H., Porwollik, S., Mottaz-Brewer, H., Petritis, B. O., Jaitly, N., Adkins, J. N., McClelland, M., Heffron, F. & Smith, R. D. (2009). Global systems-level analysis of Hfq and SmpB deletion mutants in *Salmonella*: implications for virulence and global protein translation. *PLoS ONE*, 4(3), e4809.

Antal, M., Bordeau, V., Douchin, V. & Felden, B. (2005). A small bacterial RNA regulates a putative ABC transporter. *J. Biol. Chem.*, 280(9), 7901–7908.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M. & Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, 442(7099), 203–207.

Argaman, L. & Altuvia, S. (2000). *fhlA* repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. *J. Mol. Biol.*, 300(5), 1101–1112.

Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E. G., Margalit, H. & Altuvia, S. (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.*, 11(12), 941–950.

Arluison, V., Hohng, S., Roy, R., Pellegrini, O., Régnier, P. & Ha, T. (2007). Spectroscopic observation of RNA chaperone activities of Hfq in post-transcriptional regulation by a small non-coding RNA. *Nucleic Acids Res.*, 35(3), 999–1006.

Arluison, V., Mutyam, S. K., Mura, C., Marco, S. & Sukhodolets, M. V. (2007). Sm-like protein Hfq: location of the ATP-binding site and the effect of ATP on Hfq–RNA complexes. *Protein Sci.*, 16(9), 1830–1841.

Arraiano, C. M. & Maquat, L. E. (2003). Post-transcriptional control of gene expression: effectors of mRNA decay. *Mol. Microbiol.*, 49(1), 267–276.

Axmann, I. M., Kensche, P., Vogel, J., Kohl, S., Herzel, H. & Hess, W. R. (2005). Identification of cyanobacterial non-coding RNAs by comparative genome analysis. *Genome Biol*, 6(9), R73.

Azam, T. A., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.*, 181(20), 6361–6370.

Babitzke, P. & Romeo, T. (2007). CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr. Opin. Microbiol.*, 10(2), 156–163.

Bailey, T. L. & Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, 14(1), 48–54.

Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.*, 34(Web Server issue), W369–W373.

Bang, I., Frye, J. G., McClelland, M., Velayudhan, J. & Fang, F. C. (2005). Alternative sigma factor interactions in *Salmonella*: sigma and sigma promote antioxidant defences by enhancing sigma levels. *Mol. Microbiol.*, 56(3), 811–823.

Barrick, J. E., Sudarsan, N., Weinberg, Z., Ruzzo, W. L. & Breaker, R. R. (2005). 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA*, 11(5), 774–784.

Baumeister, R., Flache, P., Melefors, O., von Gabain, A. & Hillen, W. (1991). Lack of a 5' non-coding region in Tn1721 encoded *tetR* mRNA is associated with a low efficiency of translation and a short half-life in *Escherichia coli*. *Nucleic Acids Res.*, 19(17), 4595–4600.

Beier, D., Spohn, G., Rappuoli, R. & Scarlato, V. (1997). Identification and characterization of an operon of *Helicobacter pylori* that is involved in motility and stress adaptation. *J. Bacteriol.*, 179(15), 4676–4683.

Beier, D., Spohn, G., Rappuoli, R. & Scarlato, V. (1998). Functional analysis of the *Helicobacter pylori* principal sigma subunit of RNA polymerase reveals that the spacer region is important for efficient transcription. *Mol. Microbiol.*, 30(1), 121–134.

Beitzinger, M., Peters, L., Zhu, J. Y., Kremmer, E. & Meister, G. (2007). Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.*, 4(2), 76–84.

Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.*, 340(4), 783–795.

Berezikov, E., Thuemmler, F., van Laake, L. W., Kondova, I., Bontrop, R., Cuppen, E. & Plasterk, R. H. A. (2006). Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.*, 38(12), 1375–1377.

Bernhart, S. H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P. F. & Hofacker, I. L. (2006). Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.*, 1(1), 3.

Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409(6818), 363–366.

Beuzón, C. R., Marqués, S. & Casadesús, J. (1999). Repression of IS200 transposase synthesis by RNA secondary structures. *Nucleic Acids Res.*, 27(18), 3690–3695.

Beyer, D., Skripkin, E., Wadzack, J. & Nierhaus, K. H. (1994). How the ribosome moves along the mRNA during protein synthesis. *J. Biol. Chem.*, 269(48), 30713–30717.

Blaser, M. J. (1998). *Helicobacter pylori* and gastric diseases. *BMJ*, 316(7143), 1507–1510.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331), 1453–1474.

Blomberg, P., Wagner, E. G. & Nordström, K. (1990). Control of replication of plasmid R1: the duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically *in vivo* and *in vitro* by RNase III. *EMBO J.*, 9(7), 2331–2340.

Bohn, C., Rigoulay, C. & Bouloc, P. (2007). No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. *BMC Microbiol.*, 7, 10.

Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., Chevalier, C., Helfer, A. C., Benito, Y., Jacquier, A., Gaspin, C., Vandenesch, F. & Romby, P. (2007). *Staphylococcus aureus* RNA III coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev.*, 21(11), 1353–1366.

Boneca, I. G., de Reuse, H., Epinat, J., Pupin, M., Labigne, A. & Moszer, I. (2003). A revised annotation and comparative analysis of *Helicobacter pylori* genomes. *Nucleic Acids Res.*, 31(6), 1704–1714.

Boni, I. V., Isaeva, D. M., Musychenko, M. L. & Tzareva, N. V. (1991). Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Res.*, 19(1), 155–162.

Bossi, L. & Figueroa-Bossi, N. (2007). A small RNA downregulates LamB maltoporin in *Salmonella*. *Mol. Microbiol.*, 65(3), 799–810.

Bossi, L., Maloriol, D. & Figueroa-Bossi, N. (2008). Porin biogenesis activates the sigma(E) response in *Salmonella hfq* mutants. *Biochimie*, 90(10), 1539–1544.

Bouvier, M., Sharma, C. M., Mika, F., .Nierhaus, K. H & Vogel, J. (2008). Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol. Cell*, 32(6), 827–837.

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.*, 10(2), 102–109.

Brennan, R. G. & Link, T. M. (2007). Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.*, 10(2), 125–133.

Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. (2005). Principles of microRNA-target recognition. *PLoS Biol.*, 3(3), e85.

Brock, J. E., Pourshahian, S., Giliberti, J., Limbach, P. A. & Janssen, G. R. (2008). Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5'-terminal AUG. *RNA*, 14(10), 2159–2169.

Brown, J. W. (1999). The Ribonuclease P Database. *Nucleic Acids Res.*, 27(1), 314.

Brown, L. & Elliott, T. (1996). Efficient translation of the RpoS sigma factor in *Salmonella typhimurium* requires host factor I, an RNA-binding protein encoded by the *hfq* gene. *J. Bacteriol.*, 178(13), 3763–3770.

Brownlee, G. G. (1971). Sequence of 6S RNA of *E. coli*. *Nat. New Biol.*, 229(5), 147–149.

Busch, A., Richter, A. S. & Backofen, R. (2008). IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24(24), 2849–2856.

Carpousis, A. J. (2002). The *Escherichia coli* RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. *Biochem. Soc. Trans.*, 30(2), 150–155.

Carpousis, A. J., Houwe, G. Van, Ehretsmann, C. & Krisch, H. M. (1994). Copurification of *E. coli* RNase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell*, 76(5), 889–900.

Carpousis, A. J., Vanzo, N. F. & Raynal, L. C. (1999). mRNA degradation. A tale of poly(A) and multiprotein machines. *Trends Genet.*, 15(1), 24–28.

Carter, R. J., Dubchak, I. & Holbrook, S. R. (2001). A computational approach to identify genes for functional RNAs in genomic sequences. *Nucleic Acids Res.*, 29(19), 3928–3938.

Cavanagh, A. T., Klocko, A. D., Liu, X. & Wassarman, K. M. (2008). Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of σ^{70}. *Mol. Microbiol.*, 67(6), 1242–1256.

Chant, E. L. & Summers, D. K. (2007). Indole signalling contributes to the stable maintenance of Escherichia coli multicopy plasmids. *Mol. Microbiol.*, 63(1), 35–43.

Chen, S., Lesnik, E. A., Hall, T. A., Sampath, R., Griffey, R. H., Ecker, D. J. & Blyn, L. B. (2002). A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems*, 65(2-3), 157–177.

Chen, S., Zhang, A., Blyn, L. B. & Storz, G. (2004). MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J. Bacteriol.*, 186(20), 6689–6697.

Chevalier, C., Geissmann, T., Helfer, A. & Romby, P. (2009). Probing mRNA Structure and sRNA-mRNA Interactions in Bacteria Using Enzymes and Lead(II). *Methods Mol. Biol.*, 540, 215–232.

Christiansen, J. K., Nielsen, J. S., Ebersbach., T., Valentin-Hansen, P., Søgaard-Andersen, L. & Kallipolitis, B. H. (2006). Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. *RNA*, 12(7), 1383–1396.

Cloonan, N., Forrest, A. R. R., Kolle, G., Gardiner, B. B. A., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., Robertson, A. J., Perkins, A. C., Bruce, S. J., Lee, C. C., Ranade, S. S., Peckham, H. E., Manning, J. M., McKernan, K. J. & .Grimmond, S. M (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods*, 5(7), 613–619.

Collins, J. A., Irnov, I., Baker, S. & Winkler, W. C. (2007). Mechanism of mRNA destabilization by the *glmS* ribozyme. *Genes Dev.*, 21(24), 3356–3368.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*, 16(22), 10881–10890.

Cosloy, S. D. (1973). D-serine transport system in *Escherichia coli* K-12. *J. Bacteriol.*, 114(2), 679–684.

Cromie, M. J., Shi, Y., Latifi, T. & Groisman, E. A. (2006). An RNA sensor for intracellular Mg^{2+}. *Cell*, 125(1), 71–84.

Darfeuille, F., Unoson, C., Vogel, J. & Wagner, E. G. H. (2007). An antisense RNA inhibits translation by competing with standby ribosomes. *Mol. Cell*, 26(3), 381–392.

Davis, B. M., Quinones, M., Pratt, J., Ding, Y. & Waldor, M. K. (2005). Characterization of the small untranslated RNA RyhB and its regulon in *Vibrio cholerae*. *J. Bacteriol.*, 187(12), 4005–4014.

de Hoon, M. J. L., Makita, Y., Nakai, K. & Miyano, S. (2005). Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS Comput. Biol.*, 1(3), e25.

De Lay, N. & Gottesman, S. (2009). The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *J. Bacteriol.*, 191(2), 461–476.

Deana, A. & Belasco, J. G. (2005). Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev.*, 19(21), 2526–2533.

Deiwick, J., Nikolaus, T., Erdogan, S. & Hensel, M. (1999). Environmental regulation of *Salmonella* pathogenicity island 2 gene expression. *Mol. Microbiol.*, 31(6), 1759–1773.

del Val, C., Rivas, E., Torres-Quesada, O., N.Toro & Jiménez-Zurdo, J. I. (2007). Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. *Mol. Microbiol.*, 66(5), 1080–1091.

Delany, I., G.Spohn, Pacheco, A. F., Ieva, R., Alaimo, C., Rappuoli, R. & Scarlato, V. (2002). Autoregulation of *Helicobacter pylori* Fur revealed by functional analysis of the iron-binding site. *Mol. Microbiol.*, 46(4), 1107–1122.

Delany, I., Pacheco, A. B., Spohn, G., Rappuoli, R. & Scarlato, V. (2001). Iron-dependent transcription of the *frpB* gene of *Helicobacter pylori* is controlled by the Fur repressor protein. *J. Bacteriol.*, 183(16), 4932–4937.

Delany, I., Spohn, G., Rappuoli, R. & Scarlato, V. (2001). The Fur repressor controls transcription of iron-activated and -repressed genes in *Helicobacter pylori*. *Mol. Microbiol.*, 42(5), 1297–1309.

Delany, I., Spohn, G., Rappuoli, R. & Scarlato, V. (2002). Growth phase-dependent regulation of target gene promoters for binding of the essential orphan response regulator HP1043 of *Helicobacter pylori*. *J. Bacteriol.*, 184(17), 4800–4810.

Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.*, 27(23), 4636–4641.

Dühring, U., Axmann, I. M., Hess, W. R. & Wilde, A. (2006). An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proc. Nat. Acad. Sci. U.S.A.*, 103(18), 7054–7058.

Dietz, P., Gerlach, G. & Beier, D. (2002). Identification of target genes regulated by the two-component system HP166-HP165 of *Helicobacter pylori*. *J. Bacteriol.*, 184(2), 350–362.

Ding, Y., Davis, B. M. & Waldor, M. K. (2004). Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol. Microbiol.*, 53(1), 345–354.

Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.*, 2(5), 414–424.

Doherty, N. S., Littman, B. H., Reilly, K., Swindell, A. C., Buss, J. M. & Anderson, N. L. (1998). Analysis of changes in acute-phase plasma proteins in an acute inflammatory response and in rheumatoid arthritis using two-dimensional gel electrophoresis. *Electrophoresis*, 19(2), 355–363.

Douchin, V., Bohn, C. & Bouloc, P. (2006). Down-regulation of porins by a small RNA bypasses the essentiality of the regulated intramembrane proteolysis protease RseP in Escherichia coli. *J. Biol. Chem.*, 281(18), 12253–12259.

Dsouza, M., Larsen, N. & Overbeek, R. (1997). Searching for patterns in genomic data. *Trends Genet.*, 13(12), 497–498.

Dunn, B. E., Vakil, N. B., Schneider, B. G., Miller, M. M., Zitzer, J. B., Peutz, T. & Phadnis, S. H. (1997). Localization of *Helicobacter pylori* urease and heat shock protein in human gastric biopsies. *Infect. Immun.*, 65(4), 1181–1188.

Ebeling, S., Kündig, C. & Hennecke, H. (1991). Discovery of a rhizobial RNA that is essential for symbiotic root nodule development. *J. Bacteriol.*, 173(20), 6373–6382.

Elbashir, S. M., Lendeckel, W. & Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, 15(2), 188–200.

Ender, C., Krek, A., Friedländer, M. R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N. & Meister, G. (2008). A human snoRNA with microRNA-like functions. *Mol. Cell*, 32(4), 519–528.

Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O. & Salzberg, S. L. (2000). Prediction of transcription terminators in bacterial genomes. *J. Mol. Biol.*, 301(1), 27–33.

Ernst, F. D., Stoof, J., Horrevoets, W. M., Kuipers, E. J., Kusters, J. G. & van Vliet, A. H. M. (2006). NikR mediates nickel-responsive transcriptional repression of the *Helicobacter pylori* outer membrane proteins FecA3 (HP1400) and FrpB4 (HP1512). *Infect. Immun.*, 74(12), 6821–6828.

Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Bréchemmier-Baey, D. & Putzer, H. (2005). Ribonucleases J1 and J2: two novel endoribonucleases in *B. subtilis* with functional homology to *E. coli* RNase E. *Nucleic Acids Res.*, 33(7), 2141–2152.

Fang, F. C., Libby, S. J., Buchmeier, N. A., Loewen, P. C., Switala, J., Harwood, J. & Guiney, D. G. (1992). The alternative sigma factor *katF (rpoS)* regulates *Salmonella* virulence. *Proc. Nat. Acad. Sci. U.S.A.*, 89(24), 11978–11982.

Fantappiè, L., Metruccio, M. M. E., Seib, K. L., Oriente, F., Cartocci, E., Ferlicca, F., Giuliani, M. M., Scarlato, V. & Delany, I. (2009). The RNA chaperone Hfq is involved in the stress response and virulence in *Neisseria meningitidis* and is a pleiotropic regulator of protein expression. *Infect. Immun.*

Figueroa-Bossi, N., Lemire, S., Maloriol, D., Balbontín, R., Casadesús, J. & Bossi, L. (2006). Loss of Hfq activates the σ^E-dependent envelope stress response in *Salmonella enterica*. *Mol. Microbiol.*, 62(3), 838–852.

Folichon, M., Allemand, F., Régnier, P. & Hajnsdorf, E. (2005). Stimulation of poly(A) synthesis by *Escherichia coli* poly(A)polymerase I is correlated with Hfq binding to poly(A) tails. *FEBS J.*, 272(2), 454–463.

Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Régnier, P. & Hajnsdorf, E. (2003). The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res.*, 31(24), 7302–7310.

Forsyth, M. H., Cao, P., Garcia, P. P., Hall, J. D. & Cover, T. L. (2002). Genome-wide transcriptional profiling in a histidine kinase mutant of *Helicobacter pylori* identifies members of a regulon. *J. Bacteriol.*, 184(16), 4630–4635.

Fortune, D. R., Suyemoto, M. & Altier, C. (2006). Identification of CsrC and characterization of its role in epithelial cell invasion in *Salmonella enterica* serovar Typhimurium. *Infect. Immun.*, 74(1), 331–339.

Fozo, E. M., Hemm, M. R. & Storz, G. (2008). Small toxic proteins and the antisense RNAs that repress them. *Microbiol. Mol. Biol. Rev.*, 72(4), 579–89.

Fozo, E. M., Kawano, M., Fontaine, F., Kaya, Y., Mendieta, K. S., Jones, K. L., Ocampo, A., Rudd, K. E. & Storz, G. (2008). Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol. Microbiol.*, 70(5), 1076–1093.

Franze de Fernandez, M. T., Eoyang, L. & August, J. T. (1968). Factor fraction required for the synthesis of bacteriophage Qβ-RNA. *Nature*, 219(5154), 588–590.

Franze de Fernandez, M. T., Hayward, W. S. & August, J. T. (1972). Bacterial proteins required for replication of phage Q ribonucleic acid. Pruification and properties of host factor I, a ribonucleic acid-binding protein. *J. Biol. Chem.*, 247(3), 824–831.

Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W. & Delong, E. F. (2008). Microbial community gene expression in ocean surface waters. *Proc. Nat. Acad. Sci. U.S.A.*, 105(10), 3805–3810.

Fujita, Y., Yamaguchi, K., Kamegaya, T., Sato, H., Semura, K., Mutoh, K., Kashimoto, T., Ohori, H. & Mukai, T. (2005). A novel mechanism of autolysis in *Helicobacter pylori*: possible involvement of peptidergic substances. *Helicobacter*, 10(6), 567–576.

Geisinger, E., Adhikari, R. P., Jin, R., Ross, H. F. & Novick, R. P. (2006). Inhibition of rot translation by RNAIII, a key feature of *agr* function. *Mol. Microbiol.*, 61(4), 1038–1048.

Geissmann, T. A. & Touati, D. (2004). Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J.*, 23(2), 396–405.

Gerdes, K., Gultyaev, A. P., Franch, T., Pedersen, K. & Mikkelsen, N. D. (1997). Antisense RNA-regulated programmed cell death. *Annu. Rev. Genet.*, 31, 1–31.

Gerdes, K., Thisted, T. & Martinussen, J. (1990). Mechanism of post-segregational killing by the *hok/sok* system of plasmid R1: *sok* antisense RNA regulates formation of a *hok* mRNA species correlated with killing of plasmid-free cells. *Mol. Microbiol.*, 4(11), 1807–1818.

Gildehaus, N., Neusser, T., Wurm, R. & Wagner, R. (2007). Studies on the function of the riboregulator 6S RNA from *E. coli*: RNA polymerase binding, inhibition of in vitro transcription and synthesis of RNA-directed de novo transcripts. *Nucleic Acids Res.*, 35(6), 1885–1896.

Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, 442(7099), 199–202.

Gottesman, S. (2004). The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu. Rev. Microbiol.*, 58, 303–328.

Gressmann, H., Linz, B., Ghai, R., Pleissner, K., Schlapbach, R., Yamaoka, Y., Kraft, C., Suerbaum, S., Meyer, T. F. & Achtman, M. (2005). Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet.*, 1(4), e43.

Griffin, B. (1971). Separation of ^{32}P-labelled ribonucleic acid components. The use of polyethylenimine-cellulose (TLC) as a second dimension in separating oligoribonucleotides of 4.5 S and 5 S from *E. coli*. *FEBS Lett.*, 15(3), 165–168.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R. & Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.*, 33(Database issue), D121–D124.

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P. & Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, 27(1), 91–105.

Grundy, F. J. & Henkin, T. M. (2006). From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit. Rev. Biochem. Mol. Biol.*, 41(6), 329–338.

Guillier, M. & Gottesman, S. (2006). Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. *Mol. Microbiol.*, 59(1), 231–247.

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.

Guillier, M., Gottesman, S. & Storz, G. (2006). Modulating the outer membrane with small RNAs. *Genes Dev.*, 20(17), 2338–2348.

Guisbert, E., Rhodius, V. A., Ahuja, N., Witkin, E. & Gross, C. A. (2007). Hfq modulates the σ^{E}-mediated envelope stress response and the σ^{32}-mediated cytoplasmic stress response in *Escherichia coli*. *J. Bacteriol.*, 189(5), 1963–1973.

Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, 177(14), 4121–4130.

Hajnsdorf, E., Braun, F., Haugel-Nielsen, J. & Régnier, P. (1995). Polyadenylylation destabilizes the *rpsO* mRNA of *Escherichia coli*. *Proc. Nat. Acad. Sci. U.S.A.*, 92(9), 3973–3977.

Hajnsdorf, E. & Régnier, P. (2000). Host factor Hfq of *Escherichia coli* stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proc. Nat. Acad. Sci. U.S.A.*, 97(4), 1501–1505.

Halfmann, A., Kovács, M., Hakenbeck, R. & Brückner, R. (2007). Identification of the genes directly controlled by the response regulator CiaR in *Streptococcus pneumoniae*: five out of 15 promoters drive expression of small non-coding RNAs. *Mol. Microbiol.*, 66(1), 110–126.

Haney, S. A., Platko, J. V., Oxender, D. L. & Calvo, J. M. (1992). Lrp, a leucine-responsive protein, regulates branched-chain amino acid transport genes in *Escherichia coli*. *J. Bacteriol.*, 174(1), 108–115.

Hartz, D., McPheeters, D. S., Traut, R. & Gold, L. (1988). Extension inhibition analysis of translation initiation complexes. *Methods Enzymol*, 164, 419–425.

Heeb, S., Blumer, C. & Haas, D. (2002). Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens* CHA0. *J. Bacteriol.*, 184(4), 1046–1056.

Heidrich, N., Moll, I. & Brantl, S. (2007). *In vitro* analysis of the interaction between the small RNA SR1 and its primary target *ahrC* mRNA. *Nucleic Acids Res.*, 35(13), 4331–4346.

Hemm, M. R., Paul, B. J., Schneider, T. D., Storz, G. & Rudd, K. E. (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol. Microbiol.*, 70(6), 1487–1501.

Hershberg, R., Altuvia, S. & Margalit, H. (2003). A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res.*, 31(7), 1813–1820.

Higgins, C. F. & Hardie, M. M. (1983). Periplasmic protein associated with the oligopeptide permeases of *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol.*, 155(3), 1434–1438.

Hindley, J. (1967). Fractionation of ^{32}P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *J. Mol. Biol.*, 30(1), 125–136.

Hoffmann, S., Otto, C., Kurtz, S., Sharma, C. M., Khaitovich, P., Vogel, J., Stadler, P. F. & Hackermüller, J. (2009). Fast mapping of short sequences with mismatches, insertions and deletions using index structures. (submitted).

Hoiseth, S. K. & Stocker, B. A. (1981). Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature*, 291(5812), 238–239.

Hosie, A. H. & Poole, P. S. (2001). Bacterial ABC transporters of amino acids. *Res. Microbiol.*, 152(3-4), 259–270.

Hsu, L. M., Zagorski, J., Wang, Z. & Fournier, M. J. (1985). *Escherichia coli* 6S RNA gene is part of a dual-function transcription unit. *J. Bacteriol.*, 161(3), 1162–1170.

Hüttenhofer, A. (2005). Experimental RNomics: A Global Approach to Identify Non-coding RNAs in Model Organisms. In *Handbook of RNA Biochemistry* (Hartmann, R., Bindereif, A., Schön, A. & Westhof, E., eds), pp. 643–654.

Hüttenhofer, A., Brosius, J. & Bachellerie, J. P. (2002). RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.*, 6(6), 835–843.

Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J. P. & Brosius, J. (2001). RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, 20(11), 2943–2953.

Hüttenhofer, A. & Noller, H. F. (1994). Footprinting mRNA-ribosome complexes with chemical probes. *EMBO J.*, 13(16), 3892–3901.

Hüttenhofer, A. & Vogel, J. (2006). Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res.*, 34(2), 635–646.

Hu, Z., Zhang, A., Storz, G., Gottesman, S. & Leppla, S. H. (2006). An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res.*, 34(7), e52.

Huang, C., Wolfgang, M. C., Withey, J., Koomey, M. & Friedman, D. I. (2000). Charged tmRNA but not tmRNA-mediated proteolysis is essential for *Neisseria gonorrhoeae* viability. *EMBO J.*, 19(5), 1098–1107.

Hung, S., Baldi, P. & Hatfield, G. Wesley (2002). Global gene expression profiling in *Escherichia coli* K12. The effects of leucine-responsive regulatory protein. *J. Biol. Chem.*, 277(43), 40309–40323.

Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C., Jacquier, A., Vandenesch, F. & Romby, P. (2005). *Staphylococcus aureus* RNA III and the endoribonuclease III coordinately regulate *spa* gene expression. *EMBO J.*, 24(4), 824–835.

Igarashi, K., Saisho, T., Yuguchi, M. & Kashiwagi, K. (1997). Molecular mechanism of polyamine stimulation of the synthesis of oligopeptide-binding protein. *J. Biol. Chem.*, 272(7), 4058–4064.

Ikemura, T. & Dahlberg, J. E. (1973). Small ribonucleic acids of *Escherichia coli*. I. Characterization by polyacrylamide gel electrophoresis and fingerprint analysis. *J. Biol. Chem.*, 248(14), 5024–5032.

Ikemura, T. & Dahlberg, J. E. (1973). Small ribonucleic acids of *Escherichia coli*. II. Noncoordinate accumulation during stringent control. *J. Biol. Chem.*, 248(14), 5033–5041.

Inada, M. & Guthrie, C. (2004). Identification of Lhp1p-associated RNAs by microarray analysis in *Saccharomyces cerevisiae* reveals association with coding and noncoding RNAs. *Proc. Nat. Acad. Sci. U.S.A.*, 101(2), 434–439.

Janssen, S., Reeder, J. & Giegerich, R. (2008). Shape based indexing for faster search of RNA family databases. *BMC Bioinformatics*, 9, 131.

Janzon, L., Löfdahl, S. & Arvidson, S. (1989). Identification and nucleotide sequence of the delta-lysin gene, *hld*, adjacent to the accessory gene regulator (*agr*) of *Staphylococcus aureus*. *Mol. Gen. Genet.*, 219(3), 480–485.

Johansen, J., Eriksen, M., Kallipolitis, B. & Valentin-Hansen, P. (2008). Down-regulation of outer membrane proteins by noncoding RNAs: unraveling the cAMP-CRP- and σ^E-dependent CyaR-ompX regulatory case. *J. Mol. Biol.*, 383(1), 1–9.

Johansen, J., Rasmussen, A. Aamann, Overgaard, M. & Valentin-Hansen, P. (2006). Conserved small non-coding RNAs that belong to the σ^E regulon: role in down-regulation of outer membrane proteins. *J. Mol. Biol.*, 364(1), 1–8.

Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M. & Cossart, P. (2002). An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell*, 110(5), 551–561.

Jones, A. C., Logan, R. P., Foynes, S., Cockayne, A., Wren, B. W. & Penn, C. W. (1997). A flagellar sheath protein of *Helicobacter pylori* is identical to HpaA, a putative N-acetylneuraminyllactose-binding hemagglutinin, but is not an adhesin for AGS cells. *J. Bacteriol.*, 179(17), 5643–5647.

Julio, S. M., Heithoff, D. M. & Mahan, M. J. (2000). ssrA (tmRNA) plays a role in *Salmonella enterica* serovar Typhimurium pathogenesis. *J. Bacteriol.*, 182(6), 1558–1563.

Jungblut, P. R. & Seifert, R. (1990). Analysis by high-resolution two-dimensional electrophoresis of differentiation-dependent alterations in cytosolic protein pattern of HL-60 leukemic cells. *J. Biochem. Biophys. Methods*, 21(1), 47–58.

Kaberdina, A. C., Szaflarski, W., Nierhaus, K. H. & Moll, I. (2009). An unexpected type of ribosomes induced by kasugamycin: a look into ancestral times of protein synthesis? *Mol. Cell*, 33(2), 227–236.

Kajitani, M., Kato, A., Wada, A., Inokuchi, Y. & Ishihama, A. (1994). Regulation of the *Escherichia coli hfq* gene encoding the host factor for phage Q_β. *J. Bacteriol.*, 176(2), 531–534.

Kalamorz, F., Reichenbach, B., März, W., Rak, B. & Görke, B. (2007). Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in *Escherichia coli*. *Mol Microbiol*, 65(6), 1518–1533.

Karzai, A. W., Susskind, M. M. & Sauer, R. T. (1999). SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). *EMBO J.*, 18(13), 3793–3799.

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.

Kawano, M., Aravind, L. & Storz, G. (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.*, 64(3), 738–754.

Kawano, M., Reynolds, A. A., Miranda-Rios, J. & Storz, G. (2005). Detection of 5'- and 3'-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.*, 33(3), 1040–1050.

Kawano, M., Storz, G., Rao, B. S., Rosner, J. L. & Martin, R. G. (2005). Detection of low-level promoter activity within open reading frame sequences of *Escherichia coli*. *Nucleic Acids Res.*, 33(19), 6268–6276.

Kelly, A., Goldberg, M. D., Carroll, R. K., Danino, V., Hinton, J. C. D. & Dorman, C. J. (2004). A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica* serovar Typhimurium. *Microbiology*, 150(Pt 7), 2037–2053.

Kim, K. & Lee, Y. (2004). Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases. *Nucleic Acids Res.*, 32(20), 6057–6068.

Kingsford, C. L., Ayanbule, K. & L.Salzberg, S. (2007). Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol*, 8(2), R22.

Klose, J. & Kobalz, U. (1995). Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis*, 16(6), 1034–1059.

Komarova, A. V., Tchufistova, L. S., Dreyfus, M. & Boni, I. V. (2005). AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J. Bacteriol.*, 187(4), 1344–1349.

Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., MacMenamin, P., da Piedade, I., Gunsalus, K. C., Stoffel, M. & Rajewsky, N. (2005). Combinatorial microRNA target predictions. *Nat. Genet.*, 37(5), 495–500.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science*, 294(5543), 853–858.

Landt, S. G., Abeliuk, E., McGrath, P. T., Lesley, J. A., McAdams, H. H. & Shapiro, L. (2008). Small non-coding RNAs in *Caulobacter crescentus*. *Mol. Microbiol.*, 68(3), 600–614.

Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 294(5543), 858–862.

Lease, R. A. & Belfort, M. (2000). A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. *Proc. Nat. Acad. Sci. U.S.A.*, 97(18), 9919–9924.

Lease, R. A., Cusick, M. E. & Belfort, M. (1998). Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc. Nat. Acad. Sci. U.S.A.*, 95(21), 12456–12461.

Lease, R. A., Smith, D., McDonough, K. & Belfort, M. (2004). The small noncoding DsrA RNA is an acid resistance regulator in *Escherichia coli*. *J. Bacteriol.*, 186(18), 6179–6185.

Lease, R. A. & Woodson, S. A. (2004). Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J. Mol. Biol.*, 344(5), 1211–1223.

Lee, C. A. & Falkow, S. (1990). The ability of *Salmonella* to enter mammalian cells is affected by bacterial growth state. *Proc. Nat. Acad. Sci. U.S.A.*, 87(11), 4304–4308.

Lee, R. C. & Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, 294(5543), 862–864.

Lee, T. & Feig, A. L. (2008). The RNA binding protein Hfq interacts specifically with tRNAs. *RNA*, 14(3), 514–523.

Lee, W. K., An, Y. S., Kim, K. H., Kim, S. H., Song, J. Y., Ryu, B. D., Choi, Y. J., Yoon, Y. H., Baik, S. C., Rhee, K. H. & Cho, M. J. (1997). Construction of a *Helicobacter pylori-Escherichia coli* shuttle vector for gene transfer in *Helicobacter pylori*. *Appl. Environ. Microbiol.*, 63(12), 4866–4871.

Lenz, D. H., Miller, M. B., Zhu, J., Kulkarni, R. V. & Bassler, B. L. (2005). CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*. *Mol. Microbiol.*, 58(4), 1186–1202.

Lenz, D. H., Mok, K. C., Lilley, B. N., Kulkarni, R. V., Wingreen, N. S. & Bassler, B. L. (2004). The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell*, 118(1), 69–82.

Lesnik, E. A., Sampath, R., Levene, H. B., Henderson, T. J., McNeil, J. A. & Ecker, D. J. (2001). Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res.*, 29(17), 3583–3594.

Leying, H., Suerbaum, S., Geis, G. & Haas, R. (1992). Cloning and genetic characterization of a *Helicobacter pylori* flagellin gene. *Mol. Microbiol.*, 6(19), 2863–2874.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H. & Ecker, J. R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell*, 133(3), 523–536.

Liu, J. M., Livny, J., Lawrence, M. S., Kimball, M. D., Waldor, M. K. & Camilli, A. (2009). Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.*, 10.1093, 1–10.

Liu, M. Y., Gui, G., Wei, B., Preston, J. F., Oakford, L., Yüksel, U., Giedroc, D. P. & Romeo, T. (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.

Liu, Y., Cui, Y., Mukherjee, A. & Chatterjee, A. K. (1998). Characterization of a novel RNA regulator of *Erwinia carotovora* ssp. carotovora that controls production of extracellular enzymes and secondary metabolites. *Mol. Microbiol.*, 29(1), 219–234.

Livny, J., Brencic, A., Lory, S. & Waldor, M. K. (2006). Identification of 17 *Pseudomonas aeruginosa* sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. *Nucleic Acids Res.*, 34(12), 3484–3493.

Livny, J., Fogel, M. A., Davis, B. M. & Waldor, M. K. (2005). sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. *Nucleic Acids Res.*, 33(13), 4096–4105.

Livny, J., Teonadi, H., Livny, M. & Waldor, M. K. (2008). High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. *PLoS ONE*, 3(9), e3197.

Livny, J. & Waldor, M. K. (2007). Identification of small RNAs in diverse bacterial species. *Curr. Opin. Microbiol.*, 10(2), 96–101.

Lorenz, C., von Pelchrzim, F. & Schroeder, R. (2006). Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat. Protoc.*, 1(5), 2204–2212.

Lu, C., Tej, S. S., Luo, S., Haudenschild, C. D., Meyers, B. C. & Green, P. J. (2005). Elucidation of the small RNA component of the transcriptome. *Science*, 309(5740), 1567–1569.

Lutz, R. & Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.*, 25(6), 1203–1210.

Lynn, S. P., Gardner, J. F. & Reznikoff, W. S. (1982). Attenuation regulation in the *thr* operon of *Escherichia coli* K-12: molecular cloning and transcription of the controlling region. *J. Bacteriol.*, 152(1), 363–371.

Ma, W., Cui, Y., Liu, Y., Dumenyo, C. K., Mukherjee, A. & Chatterjee, A. K. (2001). Molecular characterization of global regulatory RNA species that control pathogenicity factors in *Erwinia amylovora* and *Erwinia herbicola* pv. gypsophilae. *J. Bacteriol.*, 183(6), 1870–1880.

Macke, T. J., Ecker, D. J., Gutell, R. R., Gautheret, D., Case, D. A. & Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res.*, 29(22), 4724–4735.

Majdalani, N., Chen, S., Murrow, J., John, K. St & Gottesman, S. (2001). Regulation of RpoS by a novel small RNA: the characterization of RprA. *Mol. Microbiol.*, 39(5), 1382–1394.

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. Nat. Acad. Sci. U.S.A.*, 95(21), 12462–12467.

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.

Majdalani, N., Vanderpool, C. K. & Gottesman, S. (2005). Bacterial small RNA regulators. *Crit. Rev. Biochem. Mol. Biol.*, 40(2), 93–113.

Maki, K., Uno, K., Morita, T. & Aiba, H. (2008). RNA, but not protein partners, is directly responsible for translational silencing by a bacterial Hfq-binding small RNA. *Proc. Nat. Acad. Sci. U.S.A.*, 105(30), 10332–10337.

Mandal, M. & Breaker, R. R. (2004). Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, 5(6), 451–463.

Mandin, P., Repoila, F., Vergassola, M., Geissmann, T. & Cossart, P. (2007). Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res.*, 35(3), 962–974.

Mao, C., Evans, C., Jensen, R. V. & Sobral, B. W. (2008). Identification of new genes in *Sinorhizobium meliloti* using the Genome Sequencer FLX system. *BMC Microbiol.*, 8, 72.

Marcus, E. A. & Scott, D. R. (2001). Cell lysis is responsible for the appearance of extracellular urease in *Helicobacter pylori*. *Helicobacter*, 6(2), 93–99.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. & Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376–380.

Mariner, P. D., Walters, R. D., Espinoza, C. A., Drullinger, L. F., Wagner, S. D., Kugel, J. F. & Goodrich, J. A. (2008). Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell*, 29(4), 499–509.

Marker, C., Zemann, A., Terhörst, T., Kiefmann, M., Kastenmayer, J. P., Green, P., Bachellerie, J. P., Brosius, J. & Hüttenhofer, A. (2002). Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr. Biol.*, 12(23), 2002–2013.

Martin-Farmer, J. & Janssen, G. R. (1999). A downstream CA repeat sequence increases translation from leadered and unleadered mRNA in *Escherichia coli*. *Mol. Microbiol.*, 31(4), 1025–1038.

Maruyama, K., Sato, N. & Ohta, N. (1999). Conservation of structure and cold-regulation of RNA-binding proteins in cyanobacteria: probable convergent evolution with eukaryotic glycine-rich RNA-binding proteins. *Nucleic Acids Res.*, 27(9), 2029–2036.

Massé, E., Escorcia, F. E. & Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev.*, 17(19), 2374–2383.

Massé, E. & Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Nat. Acad. Sci. U.S.A.*, 99(7), 4620–4625.

Massé, E., Vanderpool, C. K. & Gottesman, S. (2005). Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J. Bacteriol.*, 187(20), 6962–6971.

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M. & Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Nat. Acad. Sci. U.S.A.*, 101(19), 7287–7292.

Mattatall, N. R. & Sanderson, K. E. (1996). *Salmonella typhimurium* LT2 possesses three distinct 23S rRNA intervening sequences. *J. Bacteriol.*, 178(8), 2272–2278.

Datsenko, K. A. & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Nat. Acad. Sci. U.S.A.*, 97(12), 6640–6645.

Wassarman, K. M., 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.

McArthur, S. D., Pulvermacher, S. C. & Stauffer, G. V. (2006). The *Yersinia pestis gcvB* gene encodes two small regulatory RNA molecules. *BMC Microbiol.*, 6, 52.

McClelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. & Wilson, R. K. (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, 413(6858), 852–856.

McGowan, C. C., Necheva, A. S., Forsyth, M. H., Cover, T. L. & Blaser, M. J. (2003). Promoter analysis of *Helicobacter pylori* genes with enhanced expression at low pH. *Mol. Microbiol.*, 48(5), 1225–1239.

Mückstein, U., Tafer, H., Hackermüller, J., Bernhart, S. H., Stadler, P. F. & Hofacker, I. L (2006). Thermodynamics of RNA-RNA binding. *Bioinformatics*, 22(10), 1177–1182.

Meibom, K. L., Forslund, A., Kuoppa, K., Alkhuder, K., Dubail, I., Dupuis, M., Forsberg, A. & Charbit, A. (2009). Hfq, a novel pleiotropic regulator of virulence-associated genes in *Francisella tularensis*. *Infect. Immun.*

Mika, F., Sharma, C. M., Bouvier, M., Papenfort, K. & Vogel, J. (2009). Small RNA binding to 5' mRNA coding region inhibits translational initiation. (Submitted).

Mikulecky, P. J., Kaw, M. K., Brescia, C. C., Takach, J. C., Sledjeski, D. D. & Feig, A. L. (2004). *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, *rpoS* and poly(A) RNAs. *Nat. Struct. Mol. Biol.*, 11(12), 1206–1214.

Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. Errais, Shatalin, K., Kreneva, R. A., Perumov, D. A. & Nudler, E. (2002). Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, 111(5), 747–756.

Mitarai, N., Andersson, A. M. C., Krishna, S., Semsey, S. & Sneppen, K. (2007). Efficient degradation and expression prioritization with small RNAs. *Phys. Biol.*, 4(3), 164–171.

Mizuno, T., Chou, M. Y. & Inouye, M. (1984). A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc. Nat. Acad. Sci. U.S.A.*, 81(7), 1966–1970.

Müller, S., Förster, J. & Beier, D. (2006). Repeated sequence motifs in the *Helicobacter pylori* P_{1408} promoter do not affect its transcription. *Microbiol. Res.*, 161(3), 212–221.

Møller, T., Franch, T., P.Højrup, Keene, D. R., Bächinger, H. P., Brennan, R. G. & Valentin-Hansen, P. (2002). Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol. Cell*, 9(1), 23–30.

Mohanty, B. K., Maples, V. F. & Kushner, S. R. (2004). The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*. *Mol. Microbiol.*, 54(4), 905–920.

Moll, I., Grill, S., Gualerzi, C. O. & Bläsi, U. (2002). Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol. Microbiol.*, 43(1), 239–246.

Moll, I., Hirokawa, G., Kiel, M. C., Kaji, A. & Bläsi, U. (2004). Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res.*, 32(11), 3354–3363.

Moll, I., Leitsch, D., Steinhauser, T. & Bläsi, U. (2003). RNA chaperone activity of the Sm-like Hfq protein. *EMBO Rep.*, 4(3), 284–289.

Møller, 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.

Molnár, A., Schwach, F., Studholme, D. J., Thuenemann, E. C. & Baulcombe, D. C. (2007). miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature*, 447(7148), 1126–1129.

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.

Morita, T., Mochizuki, Y. & Aiba, H. (2006). Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. *Proc. Nat. Acad. Sci. U.S.A.*, 103(13), 4858–4863.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, 5(7), 621–628.

Muffler, A., Fischer, D. & Hengge-Aronis, R. (1996). The RNA-binding protein HF-I, known as a host factor for phage Qβ RNA replication, is essential for *rpoS* translation in *Escherichia coli*. *Genes Dev.*, 10(9), 1143–1151.

Murakawa, G. J. & Nierlich, D. P. (1989). Mapping the *lacZ* ribosome binding site by RNA footprinting. *Biochemistry*, 28(20), 8067–8072.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881), 1344–1349.

Nakao, H., Watanabe, H., Nakayama, S. & Takeda, T. (1995). *yst* gene expression in *Yersinia enterocolitica* is positively regulated by a chromosomal region that is highly homologous to *Escherichia coli* host factor 1 gene (*hfq*). *Mol. Microbiol.*, 18(5), 859–865.

Neu, H. C. & Heppel, L. A. (1965). The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. *J. Biol. Chem.*, 240(9), 3685–3692.

Nicholson, A. W. (1999). Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol. Rev.*, 23(3), 371–390.

Nielsen, J. S., Bøggild, A., Andersen, C. B. F., Nielsen, G., Boysen, A., Brodersen, D. E. & Valentin-Hansen, P. (2007). An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from *Methanococcus jannaschii*. *RNA*, 13(12), 2213–2223.

Nikulin, A., Stolboushkina, E., Perederina, A., Vassilieva, I., Blaesi, U., Moll, I., Kachalova, G., Yokoyama, S., Vassylyev, D., Garber, M. & Nikonov, S. (2005). Structure of *Pseudomonas aeruginosa* Hfq protein. *Acta Crystallogr. D. Biol. Crystallogr.*, 61(Pt 2), 141–146.

Novick, R. P., Ross, H. F., Projan, S. J., Kornblum, J., Kreiswirth, B. & Moghazeh, S. (1993). Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J.*, 12(10), 3967–3975.

Ochsner, U. A., .Wilderman, P. J, Vasil, A. I. & Vasil, M. L. (2002). GeneChip expression analysis of the iron starvation response in *Pseudomonas aeruginosa*: identification of novel pyoverdine biosynthesis genes. *Mol. Microbiol.*, 45(5), 1277–1287.

Odenbreit, S., Till, M., Hofreuter, D., Faller, G. & Haas, R. (1999). Genetic and functional characterization of the *alpAB* gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. *Mol. Microbiol.*, 31(5), 1537–1548.

Oh, J. D., Kling-Bäckhed, H., Giannakis, M., Xu, J., Fulton, R. S., Fulton, L. A., Cordum, H. S., Wang, C., Elliott, G., Edwards, J., Mardis, E. R., Engstrand, L. G. & Gordon, J. I. (2006). The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proc. Nat. Acad. Sci. U.S.A.*, 103(26), 9999–10004.

O'Hara, E. B., Chekanova, J. A., Ingle, C. A., Kushner, Z. R., Peters, E. & Kushner, S. R. (1995). Polyadenylylation helps regulate mRNA decay in *Escherichia coli*. *Proc. Nat. Acad. Sci. U.S.A.*, 92(6), 1807–1811.

Ohnishi, K., Hasegawa, A., Matsubara, K., Date, T., Okada, T. & Kiritani, K. (1988). Cloning and nucleotide sequence of the *brnQ* gene, the structural gene for a membrane-associated component of the LIV-II transport system for branched-chain amino acids in *Salmonella typhimurium*. *Jpn. J. Genet.*, 63(4), 343–357.

Opdyke, J. A., Kang, J. & Storz, G. (2004). GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J. Bacteriol.*, 186(20), 6698–6705.

Ow, M. C., Perwez, T. & Kushner, S. R. (2003). RNase G of *Escherichia coli* exhibits only limited functional overlap with its essential homologue, RNase E. *Mol. Microbiol.*, 49(3), 607–622.

Padalon-Brauch, G., Hershberg, R., Elgrably-Weiss, M., Baruch, K., Rosenshine, I., Margalit, H. & Altuvia, S. (2008). Small RNAs encoded within genetic islands of *Salmonella typhimurium* show host-induced expression and role in virulence. *Nucleic Acids Res.*, 36(6), 1913–1927.

Papenfort, K., Pfeiffer, V., F.Mika, S.Lucchini, Hinton, J. C D & Vogel, J. (2006). σ^{E}-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. *Mol. Microbiol.*, 62(6), 1674–1688.

Papenfort, K., Pfeiffer, V., Lucchini, S., Sonawane, A., Hinton, J. C. D. & Vogel, J. (2008). Systematic deletion of *Salmonella* small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol. Microbiol.*, 68(4), 890–906.

Papenfort, K. & Vogel, J. (2009). Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Res. Microbiol.*, in press.

Pesci, E. C. & Pickett, C. L. (1994). Genetic organization and enzymatic activity of a superoxide dismutase from the microaerophilic human pathogen, *Helicobacter pylori*. *Gene*, 143(1), 111–116.

Petersen, L., Larsen, T. S., Ussery, D. W., On, S. L. W. & Krogh, A. (2003). RpoD promoters in *Campylobacter jejuni* exhibit a strong periodic signal instead of a -35 box. *J. Mol. Biol.*, 326(5), 1361–1372.

Pfeiffer, V., Sittka, A., Tomer, R., Tedin, K., Brinkmann, V. & Vogel, J. (2007). A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the *Salmonella* core genome. *Mol. Microbiol.*, 66(5), 1174–1191.

Pflock, M., Bathon, M., Schär, J., Müller, S., Mollenkopf, H., Meyer, T. F. & Beier, D. (2007). The orphan response regulator HP1021 of *Helicobacter pylori* regulates transcription of a gene cluster presumably involved in acetone metabolism. *J. Bacteriol.*, 189(6), 2339–2349.

Pflock, M., Dietz, P., Schär, J. & Beier, D. (2004). Genetic evidence for histidine kinase HP165 being an acid sensor of *Helicobacter pylori*. *FEMS Microbiol. Lett.*, 234(1), 51–61.

Pflock, M., Finsterer, N., Joseph, B., Mollenkopf, H., Meyer, T. F. & Beier, D. (2006). Characterization of the ArsRS regulon of *Helicobacter pylori*, involved in acid adaptation. *J. Bacteriol.*, 188(10), 3449–3462.

Pflock, M., Kennard, S., Delany, I., Scarlato, V. & Beier, D. (2005). Acid-induced activation of the urease promoters is mediated directly by the ArsRS two-component system of *Helicobacter pylori*. *Infect. Immun.*, 73(10), 6437–6445.

Phadnis, S. H., Parlow, M. H., Levy, M., Ilver, D., Caulkins, C. M., Connors, J. B. & Dunn, B. E. (1996). Surface localization of *Helicobacter pylori* urease and a heat shock protein homolog requires bacterial autolysis. *Infect. Immun.*, 64(3), 905–912.

Pichon, C. & Felden, B. (2003). Intergenic sequence inspector: searching and identifying bacterial RNAs. *Bioinformatics*, 19(13), 1707–1709.

Pichon, C. & Felden, B. (2005). Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc. Nat. Acad. Sci. U.S.A.*, 102(40), 14249–14254.

Pichon, C. & Felden, B. (2007). Proteins that interact with bacterial small RNA regulators. *FEMS Microbiol. Rev.*, 31(5), 614–625.

Pichon, C. & Felden, B. (2008). Small RNA gene identification and mRNA target predictions in bacteria. *Bioinformatics*, 24(24), 2807–2813.

Platt, T. & Yanofsky, C. (1975). An intercistronic region and ribosome-binding site in bacterial messenger RNA. *Proc. Nat. Acad. Sci. U.S.A.*, 72(6), 2399–2403.

Pánek, J., Bobek, J., Mikulík, K., Basler, M. & Vohradský, J. (2008). Biocomputational prediction of small non-coding RNAs in *Streptomyces*. *BMC Genomics*, 9, 217.

Porwollik, S. & McClelland, M. (2003). Lateral gene transfer in *Salmonella*. *Microbes Infect.*, 5(11), 977–989.

Porwollik, S., Noonan, B. & O'Toole, P. W. (1999). Molecular characterization of a flagellar export locus of *Helicobacter pylori*. *Infect. Immun.*, 67(5), 2060–2070.

Prévost, K., Salvail, H., Desnoyers, G., Jacques, J., Phaneuf, E. & Massé, E. (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.

Pulvermacher, S. C., Stauffer, L. T. & Stauffer, G. V. (2009). Role of the sRNA GcvB in regulation of *cycA* in *Escherichia coli*. *Microbiology*, 155(Pt 1), 106–114.

Pulvermacher, S. C., Stauffer, L. T. & Stauffer, G. V. (2009). The small RNA GcvB regulates *sstT* mRNA expression in *Escherichia coli*. *J. Bacteriol.*, 191(1), 238–248.

Rajewsky, N. (2006). microRNA target predictions in animals. *Nat. Genet.*, 38 Suppl, S8–13.

Rajkowitsch, L. & Schroeder, R. (2007). Dissecting RNA chaperone activity. *RNA*, 13(12), 2053–2060.

Rasmussen, A. A., Eriksen, M., Gilany, K., Udesen, C., Franch, T., Petersen, C. & Valentin-Hansen, P. (2005). Regulation of *ompA* mRNA stability: the role of a small regulatory RNA in growth phase-dependent control. *Mol. Microbiol.*, 58(5), 1421–1429.

Reeder, J. & Giegerich, R. (2005). Consensus shapes: an alternative to the Sankoff algorithm for RNA consensus structure prediction. *Bioinformatics*, 21(17), 3516–3523.

Regalia, M., Rosenblad, M. A. & Samuelsson, T. (2002). Prediction of signal recognition particle RNA genes. *Nucleic Acids Res.*, 30(15), 3368–3377.

Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(10), 1507–1517.

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

Repoila, F. & Gottesman, S. (2001). Signal transduction cascade for regulation of RpoS: temperature regulation of DsrA. *J. Bacteriol.*, 183(13), 4012–4023.

Régnier, P. & Arraiano, C. M. (2000). Degradation of mRNA in bacteria: emergence of ubiquitous features. *Bioessays*, 22(3), 235–244.

Régnier, P. & Grunberg-Manago, M. (1990). RNase III cleavages in non-coding leaders of *Escherichia coli* transcripts control mRNA stability and genetic expression. *Biochimie*, 72(11), 825–834.

Rijk, Peter De, Wuyts, Jan & Wachter, Rupert De (2003). RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics*, 19(2), 299–300.

Rivas, E. & Eddy, S. R. (2001). Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*, 2, 8.

Rivas, E., Klein, R. J., Jones, T. A. & Eddy, S. R. (2001). Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr. Biol.*, 11(17), 1369–1373.

Robbins, J. C. & Oxender, D. L. (1973). Transport systems for alanine, serine, and glycine in *Escherichia coli* K-12. *J. Bacteriol.*, 116(1), 12–18.

Robertson, G. T. & Roop, R. M. (1999). The *Brucella abortus* host factor I (HF-I) protein contributes to stress resistance during stationary phase and is a major determinant of virulence in mice. *Mol. Microbiol.*, 34(4), 690–700.

Romeo, T. (1998). Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Mol. Microbiol.*, 29(6), 1321–1330.

Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., Ge, H. & Bartel, D. P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, 127(6), 1193–1207.

Russell, R. R. (1972). Mapping of a D-cycloserine resistance locus in *Escherichia coli* K-12. *J. Bacteriol.*, 111(2), 622–624.

Saetrom, P., Sneve, R., Kristiansen, K. I., Snøve, O., Grünfeld, T., Rognes, T. & Seeberg, E. (2005). Predicting non-coding RNA genes in *Escherichia coli* with boosted genetic programming. *Nucleic Acids Res.*, 33(10), 3263–3270.

Said, N., Rieder, R., Hurwitz, R., Deckert, J., Urlaub, H. & Vogel, J. (2009). *In vivo* expression and purification of aptamer-tagged small RNA regulators. (Submitted).

Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.*, 26(2), 544–548.

Sambrook, J. & Russell, D.W. (2001). *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 3rd edition.

Santos, J. M., Drider, D., Marujo, P. E., Lopez, P. & Arraiano, C. M. (1997). Determinant role of *E. coli* RNase III in the decay of both specific and heterologous mRNAs. *FEMS Microbiol. Lett.*, 157(1), 31–38.

Sarkar, N. (1997). Polyadenylation of mRNA in prokaryotes. *Annu. Rev. Biochem.*, 66, 173–197.

Sauter, C., Basquin, J. & Suck, D. (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res.*, 31(14), 4091–4098.

Scheibe, M., Bonin, S., Hajnsdorf, E., Betat, H. & Mörl, M. (2007). Hfq stimulates the activity of the CCA-adding enzyme. *BMC Mol. Biol.*, 8, 92.

Schmitz, A., Josenhans, C. & Suerbaum, S. (1997). Cloning and characterization of the *Helicobacter pylori flbA* gene, which codes for a membrane protein involved in coordinated expression of flagellar genes. *J. Bacteriol.*, 179(4), 987–997.

Schumacher, M. A., Pearson, R. F., Møller, T., Valentin-Hansen, P. & Brennan, R. G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J*, 21(13), 3546–3556.

Selinger, D. W., Cheung, K. J., Mei, R., Johansson, E. M., Richmond, C. S., Blattner, F. R., Lockhart, D. J. & Church, G. M. (2000). RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat. Biotechnol.*, 18(12), 1262–1268.

Sharma, A. K. & Payne, S. M. (2006). Induction of expression of *hfq* by DksA is essential for *Shigella flexneri* virulence. *Mol. Microbiol.*, 62(2), 469–479.

Sharma, C. M., Darfeuille, F., Plantinga, T. H. & 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.

Shirai, M., Fujinaga, R., Akada, J. K. & Nakazawa, T. (1999). Activation of *Helicobacter pylori ureA* promoter by a hybrid *Escherichia coli-H. pylori rpoD* gene in *E. coli*. *Gene*, 239(2), 351–359.

Silvaggi, J. M., Perkins, J. B. & Losick, R. (2005). Small untranslated RNA antitoxin in *Bacillus subtilis*. *J. Bacteriol.*, 187(19), 6641–6650.

Silvaggi, J. M., Perkins, J. B. & Losick, R. (2006). Genes for small, noncoding RNAs under sporulation control in *Bacillus subtilis*. *J. Bacteriol.*, 188(2), 532–541.

Simons, R. W. & Kleckner, N. (1983). Translational control of IS10 transposition. *Cell*, 34(2), 683–691.

Singer, B. S., Shtatland, T., Brown, D. & Gold, L. (1997). Libraries for genomic SELEX. *Nucleic Acids Res.*, 25(4), 781–786.

Sittka, A., Pfeiffer, V., Tedin, K. & J.Vogel (2007). The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol. Microbiol.*, 63(1), 193–217.

Sittka, A., Sharma, C. M., Rolle, K. & Vogel, J. (2009). Deep sequencing of *Salmonella* RNA associated with heterologous Hfq proteins *in vivo* reveals small RNAs as a major target class and identifies RNA processing phenotypes. *RNA Biology*, 6(3), 1–10.

Sittka, A., S.Lucchini, Papenfort, K., Sharma, C. M., Rolle, K., Binnewies, T. T., Hinton, J. C. D. & Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.*, 4(8), e1000163.

Sittka, A. & Vogel, J. (2008). A glimpse at the evolution of virulence control. *Cell Host Microbe*, 4(4), 310–312.

Sledjeski, D. & Gottesman, S. (1995). A small RNA acts as an antisilencer of the H-NS-silenced *rcsA* gene of *Escherichia coli*. *Proc. Nat. Acad. Sci. U.S.A.*, 92(6), 2003–2007.

Sledjeski, D. D., Gupta, A. & Gottesman, S. (1996). The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*. *EMBO J.*, 15(15), 3993–4000.

Song, M., Kim, H., Kim, E. Y., Shin, M., Lee, H. C., Hong, Y., Rhee, J. H., Yoon, H., Ryu, S., Lim, S. & Choy, H. E. (2004). ppGpp-dependent stationary phase induction of genes on *Salmonella* pathogenicity island 1. *J. Biol. Chem.*, 279(33), 34183–34190.

Song, T., Mika, F., Lindmark, B., Liu, Z., Schild, S., Bishop, A., Zhu, J., Camilli, A., Johansson, J., J.Vogel & Wai, S. N. (2008). A new *Vibrio cholerae* sRNA modulates colonization and affects release of outer membrane vesicles. *Mol. Microbiol.*, 70(1), 100–111.

Sonnleitner, E., Hagens, S., Rosenau, F., Wilhelm, S., Habel, A., Jäger, K. & Bläsi, U. (2003). Reduced virulence of a *hfq* mutant of *Pseudomonas aeruginosa* O1. *Microb. Pathog.*, 35(5), 217–228.

Sonnleitner, E., Schuster, M., Sorger-Domenigg, T., Greenberg, E. P. & U.Bläsi (2006). Hfq-dependent alterations of the transcriptome profile and effects on quorum sensing in *Pseudomonas aeruginosa*. *Mol. Microbiol.*, 59(5), 1542–1558.

Sonnleitner, E., Sorger-Domenigg, Th., Madej, M. J., Findeiss, S., Hackermüller, J., A.Hüttenhofer, Stadler, P. F., Bläsi, U. & Moll, I. (2008). Detection of small RNAs in *Pseudomonas aeruginosa* by RNomics and structure-based bioinformatic tools. *Microbiology*, 154(Pt 10), 3175–3187.

Soppa, J., Straub, J., Brenneis, Ma., Jellen-Ritter, A., Heyer, R., Fischer, S., Granzow, M., Voss, B., Hess, W. R., Tjaden, B. & Marchfelder, A. (2009). Small RNAs of the halophilic archaeon *Haloferax volcanii*. *Biochem. Soc. Trans.*, 37(Pt 1), 133–136.

Spohn, G., Beier, D., Rappuoli, R. & Scarlato, V. (1997). Transcriptional analysis of the divergent *cagAB* genes encoded by the pathogenicity island of *Helicobacter pylori*. *Mol. Microbiol.*, 26(2), 361–372.

Spohn, G. & Scarlato, V. (1999). Motility of *Helicobacter pylori* is coordinately regulated by the transcriptional activator FlgR, an NtrC homolog. *J. Bacteriol.*, 181(2), 593–599.

Spohn, G. & Scarlato, V. (1999). The autoregulatory HspR repressor protein governs chaperone gene transcription in *Helicobacter pylori*. *Mol. Microbiol.*, 34(4), 663–674.

Stauffer, L. T. & Stauffer, G. V. (2005). GcvA interacts with both the alpha and sigma subunits of RNA polymerase to activate the *Escherichia coli gcvB* gene and the *gcvTHP* operon. *FEMS Microbiol. Lett.*, 242(2), 333–338.

Östberg, Y., Bunikis, I., Bergström, S. & Johansson, J. (2004). The etiological agent of Lyme disease, *Borrelia burgdorferi*, appears to contain only a few small RNA molecules. *J. Bacteriol.*, 186(24), 8472–8477.

Steffen, P., Voss, B., Rehmsmeier, M., Reeder, J. & Giegerich, R. (2006). RNAshapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics*, 22(4), 500–503.

Steglich, C., Futschik, M. E., Lindell, D., Voss, B., Chisholm, S. W. & Hess, W. R. (2008). The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *PLoS Genet.*, 4(8), e1000173.

Steitz, J. A. & Jakes, K. (1975). How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc. Nat. Acad. Sci. U.S.A.*, 72(12), 4734–4738.

Sternberg, N. L. & Maurer, R. (1991). Bacteriophage-mediated generalized transduction in *Escherichia coli* and *Salmonella typhimurium*. *Methods Enzymol*, 204, 18–43.

Stork, M., Lorenzo, M. Di, Welch, T. J. & Crosa, J. H. (2007). Transcription termination within the iron transport-biosynthesis operon of *Vibrio anguillarum* requires an antisense RNA. *J. Bacteriol.*, 189(9), 3479–3488.

Storz, G. (2002). An expanding universe of noncoding RNAs. *Science*, 296(5571), 1260–1263.

Storz, G., Altuvia, S. & Wassarman, K. M. (2005). An abundance of RNA regulators. *Annu. Rev. Biochem.*, 74, 199–217.

Storz, G., Opdyke, J. A. & Zhang, A. (2004). Controlling mRNA stability and translation with small, noncoding RNAs. *Curr. Opin. Microbiol.*, 7(2), 140–144.

Stougaard, P., Molin, S. & Nordström, K. (1981). RNAs involved in copy-number control and incompatibility of plasmid R1. *Proc. Nat. Acad. Sci. U.S.A.*, 78(10), 6008–6012.

Straub, J., Brenneis, M., Jellen-Ritter, A., Heyer, R., Soppa, J. & Marchfelder, A. (2009). Small RNAs in haloarchaea: Identification, differential expression and biological function. *RNA Biol.*, 6(3).

Suerbaum, S., Thiberge, J. M., Kansau, I., Ferrero, R. L. & Labigne, A. (1994). *Helicobacter pylori hspA-hspB* heat-shock gene cluster: nucleotide sequence, expression, putative function and immunogenicity. *Mol. Microbiol.*, 14(5), 959–974.

Sukhodolets, M. V. & Garges, S. (2003). Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry*, 42(26), 8022–8034.

Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keeffe, S., Haas, S., Vingron, M., Lehrach, H. & Yaspo, M. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, 321(5891), 956–960.

Sun, X., Zhulin, I. & Wartell, R. M. (2002). Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res.*, 30(17), 3662–3671.

Suzuma, S., Asari, S., Bunai, K., Yoshino, K., Ando, Y., Kakeshita, H., Fujita, M., Nakamura, K. & Yamane, K. (2002). Identification and characterization of novel small RNAs in the *aspS-yrvM* intergenic region of the *Bacillus subtilis* genome. *Microbiology*, 148(Pt 8), 2591–2598.

Svoboda, P. (2007). Off-targeting and other non-specific effects of RNAi experiments in mammalian cells. *Curr. Opin. Mol. Ther.*, 9(3), 248–257.

Swiercz, J. P., Hindra, Bobek, J., Bobek, J., Haiser, H. J., Berardo, C. Di, Tjaden, B. & Elliot, M. . (2008). Small non-coding RNAs in *Streptomyces coelicolor*. *Nucleic Acids Res.*, 36(22), 7240–7251.

Tafer, H. & Hofacker, I. L. (2008). RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics*, 24(22), 2657–2663.

Tang, T., Bachellerie, J., Rozhdestvensky, T., Bortolin, M., Huber, H., Drungowski, M., Elge, T., Brosius, J. & Hüttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Nat. Acad. Sci. U.S.A.*, 99(11), 7536–7541.

Tang, T., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J. P. & Hüttenhofer, A. (2005). Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.*, 55(2), 469–481.

Tenenbaum, S. A., Lager, P. J., Carson, C. C. & Keene, J. D. (2002). Ribonomics: identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays. *Methods*, 26(2), 191–198.

Tjaden, B. (2008). TargetRNA: a tool for predicting targets of small RNA action in bacteria. *Nucleic Acids Res.*, 36(Web Server issue), W109–W113.

Tjaden, B., Goodwin, S. S., Opdyke, J. A., Guillier, M., Fu, D. X., Gottesman, S. & Storz, G. (2006). Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res.*, 34(9), 2791–2802.

Tjaden, B., Saxena, R. M., Stolyar, S., Haynor, D. R., Kolker, E. & Rosenow, C. (2002). Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res.*, 30(17), 3732–3738.

Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M. & Venter, J. C. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, 388(6642), 539–547.

Tomizawa, J., Itoh, T., Selzer, G. & Som, T. (1981). Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc. Nat. Acad. Sci. U.S.A.*, 78(3), 1421–1425.

Tramonti, A., De Canio, M. & De Biase, D. (2008). GadX/GadW-dependent regulation of the *Escherichia coli* acid fitness island: transcriptional control at the *gadY-gadW* divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. *Mol. Microbiol.*, 70(4), 965–982.

Trotochaud, A. E. & Wassarman, K. M. (2004). 6S RNA function enhances long-term cell survival. *J. Bacteriol.*, 186(15), 4978–4985.

Trotochaud, A. E. & Wassarman, K. M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. *Nat. Struct. Mol. Biol.*, 12(4), 313–319.

Tsui, H. C., Leung, H. C. & Winkler, M. E. (1994). Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. *Mol. Microbiol.*, 13(1), 35–49.

Udagawa, T., Shimizu, Y. & Ueda, T. (2004). Evidence for the translation initiation of leaderless mRNAs by the intact 70 S ribosome without its dissociation into subunits in eubacteria. *J. Biol. Chem.*, 279(10), 8539–8546.

Udekwu, K. I., Darfeuille, F., Vogel, J., Reimegård, J., Holmqvist, E. & Wagner, E. G. H. (2005). Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev.*, 19(19), 2355–2366.

Ulbrandt, N. D., Newitt, J. A. & Bernstein, H. D. (1997). The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell*, 88(2), 187–196.

Ulvé, V. M., Sevin, E. W., Chéron, A. & Barloy-Hubler, F. (2007). Identification of chromosomal alpha-proteobacterial small RNAs by comparative genome analysis and detection in *Sinorhizobium meliloti* strain 1021. *BMC Genomics*, 8, 467.

Urban, J. H., Papenfort, K., J.Thomsen, .Schmitz, R. A & Vogel, J. (2007). A conserved small RNA promotes discoordinate expression of the *glmUS* operon mRNA to activate GlmS synthesis. *J. Mol. Biol.*, 373(3), 521–528.

Urban, J. H. & Vogel, J. (2007). Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res.*, 35(3), 1018–1037.

Urban, J. H. & Vogel, J. (2008). Two seemingly homologous noncoding RNAs act hierarchically to activate *glmS* mRNA translation. *PLoS Biol*, 6(3), e64.

Urbanowski, M. L., Stauffer, L. T. & Stauffer, G. V. (2000). The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol. Microbiol.*, 37(4), 856–868.

Uzzau, S., Figueroa-Bossi, N., Rubino, S. & Bossi, L. (2001). Epitope tagging of chromosomal genes in *Salmonella*. *Proc. Nat. Acad. Sci. U.S.A.*, 98(26), 15264–15269.

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.

Valverde, C., Heeb, S., Keel, C. & Haas, D. (2003). RsmY, a small regulatory RNA, is required in concert with RsmZ for GacA-dependent expression of biocontrol traits in *Pseudomonas fluorescens* CHA0. *Mol. Microbiol.*, 50(4), 1361–1379.

Valverde, C., Lindell, M., Wagner, E. G. H. & Haas, D. (2004). A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of *Pseudomonas fluorescens*. *J. Biol. Chem.*, 279(24), 25066–25074.

Valverde, C., Livny, J., Schlüter, J., Reinkensmeier, J., Becker, A. & Parisi, G. (2008). Prediction of *Sinorhizobium meliloti* sRNA genes and experimental detection in strain 2011. *BMC Genomics*, 9, 416.

Vanderpool, C. K. & Gottesman, S. (2004). Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol. Microbiol.*, 54(4), 1076–1089.

Vanet, A. & Labigne, A. (1998). Evidence for specific secretion rather than autolysis in the release of some *Helicobacter pylori* proteins. *Infect. Immun.*, 66(3), 1023–1027.

Vanet, A., Marsan, L., Labigne, A. & Sagot, M. F. (2000). Inferring regulatory elements from a whole genome. An analysis of *Helicobacter pylori* σ^{80} family of promoter signals. *J. Mol. Biol.*, 297(2), 335–353.

Vecerek, B., Moll, I. & Bläsi, U. (2005). Translational autocontrol of the *Escherichia coli hfq* RNA chaperone gene. *RNA*, 11(6), 976–984.

Vecerek, B., Moll, I. & Bläsi, U. (2007). Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. *EMBO J.*, 26(4), 965–975.

Viegas, S. C. & Arraiano, C. M. (2008). Regulating the regulators: How ribonucleases dictate the rules in the control of small non-coding RNAs. *RNA Biol.*, 5(4), 230–243.

Viegas, S. C., Pfeiffer, V., Sittka, A., Silva, I. J., Vogel, J. & Arraiano, C. M. (2007). Characterization of the role of ribonucleases in *Salmonella* small RNA decay. *Nucleic Acids Res.*, 35(22), 7651–7664.

Vogel, J., Argaman, L., Wagner, E. G. H. & Altuvia, S. (2004). The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr. Biol.*, 14(24), 2271–2276.

Vogel, J., Bartels, V., Tang, T. H., Churakov, G., Slagter-Jäger, J. G., Hüttenhofer, A. & Wagner, E. G. H. (2003). RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, 31(22), 6435–6443.

Vogel, J. & Papenfort, K. (2006). Small non-coding RNAs and the bacterial outer membrane. *Curr. Opin. Microbiol.*, 9(6), 605–611.

Vogel, J. & Sharma, C. M. (2005). How to find small non-coding RNAs in bacteria. *Biol. Chem.*, 386(12), 1219–1238.

Vogel, J. & Wagner, E. G. H. (2007). Target identification of small noncoding RNAs in bacteria. *Curr. Opin. Microbiol.*, 10(3), 262–270.

Vytvytska, O., Moll, I., Kaberdin, V. R., von Gabain, A. & Bläsi, U. (2000). Hfq (HF1) stimulates *ompA* mRNA decay by interfering with ribosome binding. *Genes Dev.*, 14(9), 1109–1118.

Wadler, C. S. & Vanderpool, C. K. (2007). A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc. Nat. Acad. Sci. U.S.A.*, 104(51), 20454–20459.

Wagner, E.G. & Darfeuille, F. (2006). Small Regulatory RNAs in Bacteria. In *Small RNAs: Analysis and regulatory functions* (Nellen, W. & Hamann, C., eds), pp. 1–29.

Wagner, E. G. H., Altuvia, S. & Romby, P. (2002). Antisense RNAs in bacteria and their genetic elements. *Adv. Genet.*, 46, 361–398.

Waldminghaus, T., Fippinger, A., Alfsmann, J. & Narberhaus, F. (2005). RNA thermometers are common in alpha- and gamma-proteobacteria. *Biol. Chem.*, 386(12), 1279–1286.

Waldminghaus, T., Heidrich, N., Brantl, S. & Narberhaus, F. (2007). FourU: a novel type of RNA thermometer in *Salmonella*. *Mol. Microbiol.*, 65(2), 413–424.

Walz, A., Pirrotta, V. & Ineichen, K. (1976). Lambda repressor regulates the switch between PR and Prm promoters. *Nature*, 262(5570), 665–669.

Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, 10(1), 57–63.

Wang, Z. & Wang, G. (2004). APD: the Antimicrobial Peptide Database. *Nucleic Acids Res.*, 32(Database issue), D590–D592.

Wargel, R. J., Hadur, C. A. & Neuhaus, F. C. (1971). Mechanism of D-cycloserine action: transport mutants for D-alanine, D-cycloserine, and glycine. *J. Bacteriol.*, 105(3), 1028–1035.

Washietl, S. & Hofacker, I. L. (2004). Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. *J. Mol. Biol.*, 342(1), 19–30.

Washietl, S., Hofacker, I. L. & Stadler, P. F. (2005). Fast and reliable prediction of noncoding RNAs. *Proc. Nat. Acad. Sci. U.S.A.*, 102(7), 2454–2459.

Washio, T., Sasayama, J. & Tomita, M. (1998). Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Res.*, 26(23), 5456–5463.

Wassarman, K.M., 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.

Wassarman, K. M. (2007). 6S RNA: a regulator of transcription. *Mol. Microbiol.*, 65(6), 1425–1431.

Wassarman, K. M. & Saecker, R. M. (2006). Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science*, 314(5805), 1601–1603.

Wassarman, K. M. & Storz, G. (2000). 6S RNA regulates *E. coli* RNA polymerase activity. *Cell*, 101(6), 613–623.

Wassarman, K. M., Zhang, A. & Storz, G. (1999). Small RNAs in *Escherichia coli*. *Trends Microbiol.*, 7(1), 37–45.

Watanabe, T., Sugiura, M. & Sugita, M. (1997). A novel small stable RNA, 6Sa RNA, from the cyanobacterium *Synechococcus* sp. strain PCC6301. *FEBS Lett.*, 416(3), 302–306.

Waters, Lauren S & Storz, Gisela (2009). Regulatory RNAs in bacteria. *Cell*, 136(4), 615–628.

Weber, A. P. M., Weber, K. L., Carr, K., Wilkerson, C. & Ohlrogge, J. B. (2007). Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiol.*, 144(1), 32–42.

Weilbacher, T., Suzuki, K., Dubey, A. K., Wang, X., Gudapaty, S., Morozov, I., Baker, C. S., Georgellis, D., Babitzke, P. & Romeo, T. (2003). A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol. Microbiol.*, 48(3), 657–670.

Weinberg, Z., Barrick, J. E., Yao, Z., Roth, A., Kim, J. N., Gore, J., Wang, J. X., Lee, E. R., Block, K. F., Sudarsan, N., Neph, S., Tompa, M., Ruzzo, W. L. & Breaker, R. R. (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.*, 35(14), 4809–4819.

Wen, Y., Feng, J., Scott, D. R., Marcus, E. A. & Sachs, G. (2007). The HP0165-HP0166 two-component system (ArsRS) regulates acid-induced expression of HP1186 α-carbonic anhydrase in *Helicobacter pylori* by activating the pH-dependent promoter. *J. Bacteriol.*, 189(6), 2426–2434.

Wilderman, P. J., Sowa, N. A., FitzGerald, D. J., FitzGerald, P. C., Gottesman, S., Ochsner, U. A. & Vasil, M. L. (2004). Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc. Nat. Acad. Sci. U.S.A.*, 101(26), 9792–9797.

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J. & Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 453(7199), 1239–1243.

Will, S., Reiche, K., Hofacker, I. L., Stadler, P. F. & Backofen, R. (2007). Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput. Biol.*, 3(4), e65.

Williams, K. P. & Bartel, D. P. (1996). Phylogenetic analysis of tmRNA secondary structure. *RNA*, 2(12), 1306–1310.

Willins, D. A., Ryan, C. W., Platko, J. V. & Calvo, J. M. (1991). Characterization of Lrp, and *Escherichia coli* regulatory protein that mediates a global response to leucine. *J. Biol. Chem.*, 266(17), 10768–10774.

Willkomm, D. K., Minnerup, J., Hüttenhofer, A. & Hartmann, R. K. (2005). Experimental RNomics in *Aquifex aeolicus*: identification of small non-coding RNAs and the putative 6S RNA homolog. *Nucleic Acids Res.*, 33(6), 1949–1960.

Wilson, J. W., Ott, C. M., zu Bentrup, K. Höner, Ramamurthy, R., Quick, L., Porwollik, S., Cheng, P., McClelland, M., Tsaprailis, G., Radabaugh, T., Hunt, A., Fernandez, D., Richter, E., Shah, M., Kilcoyne, M., Joshi, L., Nelman-Gonzalez, M., Hing, S., Parra, M., Dumars, P., Norwood, K., Bober, R., Devich, J., Ruggles, A., Goulart, C., Rupert, M., Stodieck, L., Stafford, P., Catella, L., Schurr, M. J., Buchanan, K., Morici, L., McCracken, J., Allen, P., Baker-Coleman, C., Hammond, T., Vogel, J., Nelson, R., Pierson, D. L., Stefanyshyn-Piper, H. M. & Nickerson, C. A. (2007). Space flight alters bacterial gene expression and virulence and reveals a role for global regulator Hfq. *Proc. Nat. Acad. Sci. U.S.A.*, 104(41), 16299–16304.

Wilusz, C. J. & Wilusz, J. (2005). Eukaryotic Lsm proteins: lessons from bacteria. *Nat. Struct. Mol. Biol.*, 12(12), 1031–1036.

Windbichler, N., von Pelchrzim, F., Mayer, O., Csaszar, E. & Schroeder, R. (2008). Isolation of small RNA-binding proteins from *E. coli*: evidence for frequent interaction of RNAs with RNA polymerase. *RNA Biol.*, 5(1), 30–40.

Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428(6980), 281–286.

Xiao, B., Li, W., Guo, G., Li, B., Liu, Z., Jia, K., Guo, Y., Mao, X. & Zou, Q. (2009). Identification of small noncoding RNAs in *Helicobacter pylori* by a bioinformatics-based approach. *Curr. Microbiol.*, 58(3), 258–263.

Xiao, B., Li, W., Guo, G., Li, B., Liu, Z., Tang, B., Mao, X. & Zou, Q. (2009). Screening and identification of natural antisense transcripts in *Helicobacter pylori* by a novel approach based on RNase I protection assay. *Mol. Biol. Rep.*, 36(7), 1853–1858.

Xu, F. & Cohen, S. N. (1995). RNA degradation in *Escherichia coli* regulated by 3' adenylation and 5' phosphorylation. *Nature*, 374(6518), 180–183.

Yachie, N., Numata, K., Saito, R., Kanai, A. & Tomita, M. (2006). Prediction of non-coding and antisense RNA genes in *Escherichia coli* with Gapped Markov Model. *Gene*, 372, 171–181.

Yamanaka, K., Fang, L. & Inouye, M. (1998). The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. *Mol. Microbiol.*, 27(2), 247–255.

Yassour, M., Kaplan, T., Fraser, H. B., Levin, J. Z., Pfiffner, J., Adiconis, X., Schroth, G., Luo, S., Khrebtukova, I., Gnirke, A., Nusbaum, C., Thompson, D., Friedman, N. & Regev, A. (2009). *Ab initio* construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *Proc. Nat. Acad. Sci. U.S.A.*, 106(9), 3264–3269.

Yoder-Himes, D. R., Chain, P. S G, Zhu, Y., Wurtzel, O., Rubin, E. M., Tiedje, James M & Sorek, R. (2009). Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc. Nat. Acad. Sci. U.S.A.*, 106(10), 3976–3981.

Yuan, G., Klämbt, C., Bachellerie, J., Brosius, J. & Hüttenhofer, A. (2003). RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.*, 31(10), 2495–2507.

Yusupova, G., Jenner, L., Rees, B., Moras, D. & Yusupov, M. (2006). Structural basis for messenger RNA movement on the ribosome. *Nature*, 444(7117), 391–394.

Yusupova, G. Z., Yusupov, M. M., Cate, J. H. & Noller, H. F. (2001). The path of messenger RNA through the ribosome. *Cell*, 106(2), 233–241.

Zhang, A., Altuvia, S., Tiwari, A., Argaman, L., Hengge-Aronis, R. & Storz, G. (1998). The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J.*, 17(20), 6061–6068.

Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C. & Storz, G. (2002). The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell*, 9(1), 11–22.

Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G. & Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.*, 50(4), 1111–1124.

Zhang, Y., Sun, S., Wu, T., Wang, J., Liu, C., Chen, L., Zhu, X., Zhao, Y., Zhang, Z., Shi, B., Lu, H. & Chen, R. (2006). Identifying Hfq-binding small RNA targets in *Escherichia coli*. *Biochem. Biophys. Res. Commun.*, 343(3), 950–955.

Zhang, Y., Zhang, Z., Ling, L., Shi, B. & Chen, R. (2004). Conservation analysis of small RNA genes in *Escherichia coli*. *Bioinformatics*, 20(5), 599–603.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G. J., Wang, X. & Qi, Y. (2007). A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev.*, 21(10), 1190–1203.

Ziolkowska, K., Derreumaux, P., Folichon, M., Pellegrini, O., Régnier, P., Boni, I. V. & Hajnsdorf, E. (2006). Hfq variant with altered RNA binding functions. *Nucleic Acids Res.*, 34(2), 709–720.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31(13), 3406–3415.

CHAPTER 10

APPENDICES

The following Table 10.1 lists the bacterial strains that were used in Chapters 3, 4, and 5 of this thesis.

Table 10.1: Bacterial strains.

Strain	Relevant markers/ genotype	Reference/ source
S. typhimurium		
SL1344	Str^R *hisG rpsL xyl*	Hoiseth & Stocker, 1981, provided by D. Bumann, MPI-IB Berlin
JVS-0255	SL1344 Δ*hfq::Cm*^{R}	Sittka *et al.*, 2007
JVS-0236	SL1344 Δ*gcvB::Km*^{R}	this study
JVS-0617	SL1344 Δ*gcvB*/Δ*hfq*	this study
JVS-1338	SL1344 *hfq*$^{\text{Flag}}$	Pfeiffer *et al.*, 2007
E. coli		
TOP10	*mcrA* Δ*(mrr-hsdRMS-mcrBC)* Φ80*lacZ*ΔM15 Δ*lacX*74 *deoR recA1 araD139* Δ*(ara-leu)*7697 *galU galK rpsL endA1 nupG*	Invitrogen
TOP10 F'	F´*lacIq* Tn*10* (Tet^R) *mcrA* Δ*(mrr-hsdRMS-mcrBC)* Φ80*lacZ* ΔM15 Δ*lacX*74 *deoR recA1 araD*139 Δ*(ara-leu)*7697 *galU galK rpsL endA1 nupG*	Invitrogen
JVS-6081	TOP10 Δ*gcvB::Km*^{R}	this study

Plasmids that were used or constructed in Chapters 3 and 4 of this thesis are listed in the following Table 10.2.

Table 10.2: Plasmids that were used or constructed in Chapters 3 and 4 of this thesis.

Name	Synonym	Relevant fragment	Comment	Origin/ marker	Reference
pKD3			Template for mutant construction; carries chloramphenicol cassette	oriRγ/ Amp^R	Datsenko & Wanner, 2000
pKD4			Template for mutant construction; carries kanamycin cassette	oriRγ/ Amp^R	Datsenko & Wanner, 2000
pKD46		P$_{araB}$-γ-β-*exo*	Temperature sensitive λ red recombinase expression plasmid	oriR101/ Amp^R	Datsenko & Wanner, 2000
pCP20			Temperature sensitive FLP recombinase expression plasmid	oriR101/ Amp^R, Cm^R	Datsenko & Wanner, 2000
pJV300			ColE1 control plasmid, based on pZE12-*luc*, -1 site of P$_{LlacO}$ promoter religated to second position of XbaI site (destroyed), yields $\approx$ 50 nt nonsense transcript derived from *rrnB* terminator	ColE1/ Amp^R	Sittka *et al.*, 2007
pJV968-1		'*lacZ*'	ColE control plasmid, carries 1.5 kb internal *lacZ* fragment	ColE1/ Amp^R	Vogel *et al.*, 2004
pJV846-11		P$_{LlacO}$ *gcvB*	*Salmonella* SL1344 *gcvB* high-copy expression plasmid, *gcvB* is controlled by the constitutive P$_{LlacO}$ promoter	ColE1/ Amp^R	Sharma *et al.*, 2007
pZE12-*luc*		*luc*	General expression plasmid	ColE1/ Amp^R	Lutz & Bujard, 1997
pTP11	control plasmid		Control plasmid based on pJV300, ColE1 origin replaced by p15A origin	p15A/ Amp^R	Sharma *et al.*, 2007
pTP24	control plasmid	'*lacZ*'	Low copy control plasmid, based on pJV968-1; ColE1 origin replaced by p15A origin	p15A/ Amp^R	Sharma *et al.*, 2007
pTP02		P$_{gcvB}$-*gcvB*	*Salmonella* SL1344 *gcvB* high-copy expression plasmid, *gcvB* is controlled by its own promoter	ColE1/ Amp^R	Sharma *et al.*, 2007
pTP05	p*gcvB*	P$_{gcvB}$-*gcvB*	*Salmonella* SL1344 *gcvB* mid-copy expression plasmid, *gcvB* is controlled by its own promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pTP09	pP$_L$ *gcvB*	P$_{LlacO}$-*gcvB*	*Salmonella* SL1344 *gcvB* mid-copy expression plasmid, *gcvB* is controlled by the constitutive P$_{LlacO}$ promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pJL03-15	p*gcvB*$_{\Delta R1}$	P$_{wt}$-*gcvB*$_{\Delta R1}$	*Salmonella* SL1344 *gcvB*$_{\Delta R1}$ (deletion of position 66 - 89) mid-copy expression plasmid, *gcvB*$_{\Delta R1}$ is controlled by its own promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pJL16-10	p*gcvB*$_{\Delta R2}$	P$_{wt}$-*gcvB*$_{\Delta R2}$	*Salmonella* SL1344 *gcvB*$_{\Delta R2}$ (deletion of position 136 - 144) mid-copy expression plasmid, *gcvB*$_{\Delta R1}$ is controlled by its own promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pJL05-16	p*gcvB*$_{5'\Delta}$	P$_{wt}$-*gcvB*$_{5'\Delta}$	*Salmonella* SL1344 *gcvB*$_{5'\Delta}$ (deletion of position 1 - 91) mid-copy expression plasmid, *gcvB*$_{5'\Delta}$ is controlled by its own promoter	p15A/ Amp^R	Sharma *et al.*, 2007

continued on next page

Name	Synonym	Relevant fragment	Comment	Origin/ marker	Reference
pJL01-1	p*gcvB*$_{3'\Delta}$	P_{wt}-*gcvB*$_{3'\Delta}$	*Salmonella* SL1344 *gcvB*$_{3'\Delta}$ (deletion of position 135 - 206) mid-copy expression plasmid, *gcvB*$_{3'\Delta}$ is controlled by its own promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pJL17-6	p*gcvB*$_{3'\Delta T}$	P_{wt}-*gcvB*$_{3'\Delta T}$	*Salmonella* SL1344 *gcvB*$_{3'\Delta T}$ (deletion of bp 135 - 206 and stabilized terminator) mid-copy expression plasmid, *gcvB*$_{3'\Delta T}$ is controlled by its own promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pJL22	pP_L*gcvB*$_{\Delta R1}$	P_{LlacO}-*gcvB*$_{\Delta R1}$	*Salmonella* SL1344 *gcvB*$_{\Delta R1}$ (deletion of position 66 - 89) mid-copy expression plasmid, *gcvB*$_{\Delta R1}$ is controlled by the constitutive P_{LlacO} promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pJL23	pP_L*gcvB*$_{\Delta R2}$	P_{LlacO}-*gcvB*$_{\Delta R2}$	*Salmonella* SL1344 *gcvB*$_{\Delta R2}$ (deletion of position 136 - 144) mid-copy expression plasmid, *gcvB*$_{\Delta R2}$ is controlled by the constitutive P_{LlacO} promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pJL29-4	pP_L*gcvB*$_{5'\Delta}$	P_{LlacO} -*gcvB*$_{5'\Delta}$	*Salmonella* SL1344 *gcvB*$_{5'\Delta}$ (deletion of position 1 - 91) mid-copy expression plasmid, *gcvB*$_{5'\Delta}$ is controlled by the constitutive P_{LlacO} promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pJL13-12	p*gcvB*$_{3'\Delta T'}$	P_{wt}-*gcvB*$_{3'\Delta T'}$	*Salmonella* SL1344 *gcvB*$_{3'\Delta T}$ (deletion of bp 135-206 and mutated SL3) mid-copy expression plasmid, *gcvB*$_{3'\Delta T'}$ is controlled by its own promoter	p15A/ Amp^R	Sharma *et al.*, 2007
pJV752-1			pZE12-*luc* with modified p15A origin (no XbaI site)	p15A/ Amp^R	J. Vogel, unpublished
pXG-0 (pJU-004)	*no gfp*	*luc*	non-fluorescent control plasmid, no *gfp*	pSC101*/ Cm^R	Urban & Vogel, 2007
pXG-1 (pJV859-8)	*gfp*	P_{LtetO}-*gfp*	Fluorescent GFP control plasmid, constitutively expresses full-length *gfp*	pSC101*/ Cm^R	Urban & Vogel, 2007
pXG-10		P_{LtetO}-*lacZ::gfp*	standard plasmid for directional cloning of a target mRNA as N-translational fusion to GFP	pSC101*/ Cm^R	Urban & Vogel, 2007
pXG-20		P_{LtetO}-*lacZ::gfp*	plasmid for RACE *gfp* fusion cloning	pSC101*/ Cm^R	Urban & Vogel, 2007
pJL18-1	*dppA::gfp*	P_{LtetO}-*dppA::gfp*	*Salmonella dppA* translational GFP fusion plasmid	pSC101*/ Cm^R	Sharma *et al.*, 2007
pJL19-1	*oppA::gfp*	P_{LtetO}-*oppA::gfp*	*Salmonella oppA* translational GFP fusion plasmid	pSC101*/ Cm^R	Sharma *et al.*, 2007
pJL24-1	*gltI::gfp*	P_{LtetO}-*gltI::gfp*	*Salmonella gltI* translational GFP fusion plasmid	pSC101*/ Cm^R	Sharma *et al.*, 2007
pJL20-1	*livJ::gfp*	P_{LtetO}*livJ::gfp*	*Salmonella livJ* translational GFP fusion plasmid	pSC101*/ Cm^R	Sharma *et al.*, 2007
pJL31-24	*livK::gfp*	P_{LtetO}-*livK::gfp*	*Salmonella livK* translational GFP fusion plasmid	pSC101*/ Cm^R	Sharma *et al.*, 2007
pJL27-2	*argT::gfp*	P_{LtetO}-*argT::gfp*	*Salmonella argT* translational GFP fusion plasmid	pSC101*/ Cm^R	Sharma *et al.*, 2007
pTP28	STM4351::gfp	P_{LtetO}-STM4351*::gfp*	*Salmonella* STM4351 translational GFP fusion plasmid	pSC101*/ Cm^R	Sharma *et al.*, 2007
pJU-63	*ompR::gfp*	P_{LtetO}-*ompR::gfp*	*E. coli ompR* fused to *gfp* at aa 35	pSC101*/ Cm^R	Urban & Vogel, 2007

continued on next page

Name	Synonym	Relevant fragment	Comment	Origin/ marker	Reference
pJL45-3	*gltI*$_{\Delta CA}$*::gfp*	P_{LtetO}-*gltI*$_{\Delta CA}$*::gfp*	*Salmonella gltI::gfp* fusion plasmid with deletion of CA-rich region from 5'UTR of *gltI*	pSC101*/ Cm^R	Sharma *et al.*, 2007
pJL50-11	*ompR*$_{CA}$*::gfp*	P_{LtetO}-*ompR*$_{CA}$*::gfp*	*E. coli ompR ::gfp* with insertion of CA-rich element from *Salmonella gltI* 5'UTR	pSC101*/ Cm^R	Sharma *et al.*, 2007
pJL56-2	*gltI*$_{M2}$*::gfp*	P_{LtetO}-*gltI*$_{M2}$*::gfp*	*Salmonella gltI::gfp* fusion plasmid with two mutations to disrupt stem-loop structure in 5'UTR of *gltI*	pSC101*/ Cm^R	Sharma *et al.*, 2007
pFS133-3	*ygjU::gfp*	P_{LtetO}-*ygjU::gfp*	*Salmonella ygjU (sstT)* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pFM27-1	*yaeC::gfp*	P_{LtetO}-*yaeC::gfp*	*Salmonella yaeC* translational GFP fusion plasmid	pSC101*/ Cm^R	F. Mika, unpublished
pJL69-5	*gdhA::gfp*	P_{LtetO}-*gdhA::gfp*	*Salmonella gdhA* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pFS116-1	*asd::gfp*	P_{LtetO}-*asd::gfp*	*Salmonella asd* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pFS103-3	*lrp::gfp*	P_{LtetO}-*lrp::gfp*	*Salmonella lrp* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pJL68-1	*ilvC::gfp*	P_{LtetO}-*ilvC::gfp*	*Salmonella ilvC* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pFS121-1	*ciA::gfp*	P_{LtetO}-*iciA::gfp*	*Salmonella iciA* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pFS105-3	*brnQ::gfp*	P_{LtetO}-*brnQ::gfp*	*Salmonella brnQ* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pSP25-7	*ilvE::gfp*	P_{LtetO}-*ilvE::gfp*	*Salmonella ilvE* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pSP20-1	*thrL::gfp*	P_{LtetO}-*thrL::gfp*	*Salmonella thrL* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pSP21-2	*ybdH::gfp*	P_{LtetO}-*ybdH::gfp*	*Salmonella ybdH* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pFS115-2	*ndk::gfp*	P_{LtetO}-*ndk::gfp*	*Salmonella ndk* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pFS117-1	*serA::gfp*	P_{LtetO}-*serA::gfp*	*Salmonella serA* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pJL30-14	*cycA::gfp*	P_{LtetO}-*cycA::gfp*	*Salmonella cycA* translational GFP fusion plasmid to 19th amino acid	pSC101*/ Cm^R	this study
pJL83-2	*cycA_10th::gfp*	P_{LtetO}-*cycA10th::gfp*	*Salmonella cycA* translational GFP fusion plasmid to 10th amino acid	pSC101*/ Cm^R	this study
pJL70-9	*ydgR ::gfp*	P_{LtetO}-*ydgR::gfp*	*Salmonella ydgR (tppB)* translational GFP fusion plasmid	pSC101*/ Cm^R	this study
pKP6-21		P_{BAD}-*micA*	*Salmonella* MicA expression plasmid, *micA* is under control of P_{BAD} promoter	pBR322/ Amp^R	Papenfort *et al.*, unpublished
pKP8-35	control plasmid		pBAD control plasmid, expresses $\approx$ 50 nt nonsense RNA derived from *rrnB* terminator	pBR322/ Amp^R	Papenfort *et al.*, 2006
pKP1-1	pBAD-GcvB	P_{BAD}-*gcvB*	GcvB expression plasmid, *gcvB* is controlled by the plasmid-borne P_{BAD} promoter	pBR322/ Amp^R	this study
pKP2-6	pBAD-GcvB$_{\Delta R1}$	P_{BAD}-*gcvB*$_{\Delta R1}$	GcvB $_{\Delta R1}$ (deletion of position 66 - 89) expression plasmid, *gcvB*$_{\Delta R1}$is controlled by the plasmid-borne P_{BAD} promoter	pBR322/ Amp^R	this study

continued on next page

Name	Synonym	Relevant fragment	Comment	Origin/ marker	Reference
pKP30-1	pBAD-GcvB$_{\Delta R2}$	P$_{BAD}$-*gcvB*$_{\Delta R2}$	GcvB$_{\Delta R2}$ (deletion of position 136 - 144) expression plasmid, *gcvB*$_{\Delta R2}$ is controlled by the plasmid-borne P$_{BAD}$ promoter	pBR322/ Amp^R	this study
pJL36-5	pP$_L$-*gcvB*$_{\Delta R1\ \&\ \Delta R2}$	P$_{LlacO}$-*gcvB*$_{\Delta R1\ \&\ \Delta R2}$	*Salmonella* SL1344 *gcvB*$_{\Delta R1\ \&\ \Delta R2}$ (deletion of position 66 - 89 and 136 - 144) mid-copy expression plasmid, *gcvB*$_{\Delta R1\ \&\ \Delta R2}$ is controlled by the constitutive P$_{LlacO}$ promoter	p15A/ Amp^R	this study
pJL57-1	pP$_L$-*gcvB*$_{5'\Delta\ \&\ \Delta R2}$	P$_{LlacO}$-*gcvB*$_{5'\Delta\ \&\ \Delta R2}$	*Salmonella* SL1344 *gcvB*$_{5'\Delta\ \&\ \Delta R2}$ (deletion of position 1 - 91 and 136 - 144) mid-copy expression plasmid, *gcvB*$_{5'\Delta\ \&\ \Delta R2}$ is controlled by the constitutive P$_{LlacO}$ promoter	p15A/ Amp^R	this study
pJL65-3	pP$_L$*gcvB*$_{\Delta SL2}$	P$_{LlacO}$-*gcvB*$_{\Delta SL2}$	*Salmonella* SL1344 *gcvB*$_{\Delta SL2}$ (deletion of position 92 - 113) mid-copy expression plasmid, *gcvB*$_{\Delta SL2}$ is controlled by the constitutive P$_{LlacO}$ promoter	p15A/ Amp^R	this study
pJL66-12	pP$_L$-*gcvB*$_{\Delta SL2\ \&\ \Delta SL3}$	P$_{LlacO}$-*gcvB*$_{\Delta SL2\ \&\ \Delta SL3}$	*Salmonella* SL1344 *gcvB*$_{\Delta SL2\ \&\ \Delta SL3}$ (deletion of position 92 - 134) mid-copy expression plasmid, *gcvB*$_{\Delta SL2\ \&\ \Delta SL3}$ is controlled by the constitutive P$_{LlacO}$ promoter	p15A/ Amp^R	this study
pFS127-2	pP$_L$-*gcvB*$_{SL1,\ SL4\ \&\ SL5}$	P$_{LlacO}$-*gcvB*$_{SL1,\ SL4\ \&\ SL5}$	*Salmonella* SL1344 *gcvB*$_{SL1,\ SL4\ \&\ SL5}$ (deletion of position 66 - 144) mid-copy expression plasmid, *gcvB*$_{SL1,\ SL4\ \&\ SL5}$ is controlled by the constitutive P$_{LlacO}$ promoter	p15A/ Amp^R	this study
pFS129-2	pP$_L$*gcvB*$_{SL4\ \&\ SL5}$	P$_{LlacO}$-*gcvB*$_{SL4\ \&\ SL5}$	*Salmonella* SL1344 *gcvB*$_{SL4\ \&\ SL5}$ (deletion of position 1 - 144) mid-copy expression plasmid, *gcvB*$_{SL4\ \&\ SL5}$ is controlled by the constitutive P$_{LlacO}$promoter	p15A/ Amp^R	this study
pFS130-1	pP$_L$-*gcvB*$_{R2,\ SL4\ \&\ SL5}$	P$_{LlacO}$-*gcvB*$_{R2,\ SL4\ \&\ SL5}$	*Salmonella* SL1344 *gcvB*$_{R2,\ SL4\ \&\ SL5}$ (deletion of position 1 - 134) mid-copy expression plasmid, *gcvB*$_{R2,\ SL4\ \&\ SL5}$ is controlled by the constitutive P$_{LlacO}$ promoter	p15A/ Amp^R	this study
pFS131-1	p*gcvB*$_{\Delta R1\ \&\ 3'\Delta T}$	P$_{wt}$-*gcvB*$_{\Delta R1\ \&\ 3'\Delta T}$	*Salmonella* SL1344 *gcvB*$_{\Delta R1\ \&\ 3'\Delta T}$ (66 - 89, deletion of bp 135 - 206 and stabilized terminator) mid-copy expression plasmid, *gcvB*$_{\Delta R1\ \&\ 3'\Delta T}$ is controlled its own promoter	p15A/ Amp^R	this study
pJL73-14	pP$_L$*gcvB*$_{SL1\ \&\ SL5}$	P$_{LlacO}$-*gcvB*$_{SL1\ \&\ SL5}$	*Salmonella* SL1344 *gcvB*$_{SL1\ \&\ SL5}$ (deletion of position 66 - 177) mid-copy expression plasmid, *gcvB*$_{SL1\ \&\ SL5}$ is controlled by the constitutive P$_{LlacO}$ promoter	p15A/ Amp^R	this study

continued on next page

Name	Synonym	Relevant fragment	Comment	Origin/ marker	Reference
pJL79-16	pP_L*gcvB*$_{M2, R2, SL4 \& SL5}$	P_{LlacO}-*gcvB*$_{M2, R2, SL4 \& SL5}$	*Salmonella* SL1344 *gcvB*$_{M2, R2, SL4 \& SL5}$ (deletion of position 1- 134, $G_{143} \rightarrow C$, $C_{158} \rightarrow G$) mid-copy expression plasmid, *gcvB*$_{M2, R2, SL4 \& SL5}$ is controlled by the constitutive P_{LlacO} promoter	p15A/ Amp^R	this study
pJL78-11	pP_L*gcvB*$_{5'\Delta 12nt, SL1 \& SL5}$	P_{LlacO}-*gcvB*$_{5'\Delta 12nt, SL1 \& SL5}$	*Salmonella* SL1344 *gcvB*$_{5'\Delta 12nt, SL1 \& SL5}$ (deletion of position 1 - 12, 66 - 177) mid-copy expression plasmid, *gcvB*$_{5'\Delta 12nt, SL1 \& SL5}$ is controlled by the constitutive P_{LlacO} promoter	p15A/ Amp^R	this study
pJL77-3	p*gcvB*$_{5'\Delta 12nt, \Delta R1 \& 3'\Delta T}$	P_{wt}-*gcvB*$_{5'\Delta 12nt, \Delta R1 \& 3'\Delta T}$	*Salmonella* SL1344 *gcvB*$_{5'\Delta 12nt, \Delta R1 \& 3'\Delta T}$ (1 - 12, 66 - 89, deletion of bp 135 - 206 and stabilized terminator) mid-copy expression plasmid, *gcvB*$_{5'\Delta 12nt, \Delta R1 \& 3'\Delta T}$ is controlled its own promoter	p15A/ Amp^R	this study
pJL85-4	p*gcvB*$_{\Delta PL}$	*gcvB*$_{\Delta PL}$	*Salmonella* SL1344 *gcvB*$_{\Delta PL}$ (deletion of position (-1) - (-35)) mid-copy expression plasmid, *gcvB*$_{\Delta PL}$ has a deletion of the constitutive P_{LlacO} promoter	p15A/ Amp^R	this study
pSP9-1	pP_L-*gcvB*$_{SL1 \& SL5, C3}$	P_{LlacO}-*gcvB*$_{SL1 \& SL5, C3}$	*Salmonella* SL1344 *gcvB*$_{SL1 \& SL5, C3}$ (deletion of position 66 - 177, single nucleotide exchange $T_3 \rightarrow C$) mid-copy expression plasmid, *gcvB*$_{SL1 \& SL5, C3}$ is controlled by the constitutive P_{LlacO} promoter	p15A/ Amp^R	this study
pSP11-1	pP_L-*gcvB*$_{SL1 \& SL5, C8}$	P_{LlacO}-*gcvB*$_{SL1 \& SL5, C8}$	*Salmonella* SL1344 *gcvB*$_{SL1 \& SL5, C8}$ (deletion of position 66 - 177, single nucleotide exchange $G_8 \rightarrow C$) mid-copy expression plasmid, *gcvB*$_{SL1 \& SL5, C8}$ is controlled by the constitutive P_{LlacO} promoter	p15A/ Amp^R	this study
pSP10-1	pP_L-*gcvB*$_{SL1 \& SL5, G11}$	P_{LlacO}-*gcvB*$_{SL1 \& SL5, G11}$	*Salmonella* SL1344 *gcvB*$_{SL1 \& SL5, G11}$ (deletion of position 66 - 177, single nucleotide exchange $C_{11} \rightarrow G$) mid-copy expression plasmid, *gcvB*$_{SL1 \& SL5, G11}$ is controlled by the constitutive P_{LlacO} promoter	p15A/ Amp^R	this study

10.1. Appendix to Chapter 3

This section contains supplementary Figures and Tables to Chapter 3.

Table 10.3: DNA oligonucleotides used in this study.

Sequences are given in 5' → 3' direction and 5'P denotes a 5'-mono-phosphate.

Name	Sequence	Used for
pZE-T1	CGGCGGATTTGTCCTACT	T7 template
PLlacO-C	5' P-GTGCTCAGTATCTTGTTATCCG	sRNA cloning
pZE-B	GGCGTATCTCTTCATAGCCTTAT	sRNA cloning
pZE-A	GTGCCACCTGACGTCTAAGA	sRNA cloning
JVO-0076	GAAGTATTACAGGTTGTTGGTG	Knockout construction
JVO-0077	GCATCATAACGGTCAAACA	Knockout construction
JVO-0131	TTCTACCAGCAAATACCTATAGTGGCGGCACTTCCTGAGCC GGAAGTGTAGGCTGGAGCTGCTTC	Knockout construction
JVO-0132	TCGCGATCGCAAGGTAAAAAAAAGCACCGCAATTAGGCGGT GCTAGGTCCATATGAATATCCTCCTTAG	Knockout construction
JVO-0133	TTCTACCAGCAAATACCTATAGTGGCGGCACTTCCTGAGCC GGAAGTGTAGGCTGGAGCTGCTTC	Knockout construction
JVO-0134	TAACCGTTTGTTATACAAAAAAAAGCACCGCAATATGGCGG TGCTAGGTCCATATGAATATCCTCCTTAG	Knockout construction
JVO-0135	GTTTTTCTCGAGCGGTCAGCAGGAGTGAA	Knockout construction
JVO-0136	GTTTTTTCTAGACATCGTCTCTGACGGCA	Knockout construction
JVO-0137	GTTTTTCTCGAGCGGCGGAACAGTTTTA	Knockout construction, sRNA cloning
JVO-0138	GTTTTTTCTAGACCGATAACGATACCGGTAT	Knockout construction, sRNA cloning
JVO-0155	CCGTATGTAGCATCACCTTC	Northern blot probe
JVO-0237	ACTTCCTGAGCCGGAAC	sRNA cloning
JVO-0322	CTACGGCGTTTCACTTCTGAGTTC	Northern blot probe
JVO-0367	ACTGACATGGAGGAGGGA	GFP fusion cloning
JVO-0424	GTTTTTGCTAGCCATCCCTGACTTCTTCAAG	GFP fusion cloning
JVO-0427	GTTTTTGCTAGCAGACAGTCCCATGAC	GFP fusion cloning
JVO-0619	CTCTAGAGGCATCAAATAAAAC	sRNA cloning
JVO-0656	GTTTTGCTAGCAGTGAGTATTCCCGCTG	GFP fusion cloning
JVO-0728	GTTTTTGCTAGCGATACATCCTGCCAATAAC	GFP fusion cloning
JVO-0731	GTTTTATGCATCAGAATAGCACCCTGCG	GFP fusion cloning
JVO-0732	GTTTTTGCTAGCTAACAGCATGACAATAAGTTTT	GFP fusion cloning
JVO-0742	5' P-AAAAAAGGGTAGTCTCGCTAC	sRNA cloning, T7 template
JVO-0743	5' P-GCCGCCACTATAGGTATT	sRNA cloning
JVO-0744	5' P-ATTGGTCTGCGATTCAGA	sRNA cloning
JVO-0745	5' P-ACCGTAAGCCAAAAGC	sRNA cloning
JVO-0746	5' P-CAATTGGTCTGCGATTC	sRNA cloning
JVO-0749	TTCGTTCCGGCTCAGGA	Northern blot probe
JVO-0750	AATCACTATGGACAGACAGGGTA	Northern blot probe
JVO-0796	GTTTTTGCTAGCGCCCAGACCTATCAGCAA	GFP fusion cloning
JVO-0800	GTTTTTGCTAGCAACAATCCCTGCGATTATT	GFP fusion cloning
JVO-0892	5' P-AAAAAAAACGGTAGCGTTTCCGCTACCGTGGTCTGA	sRNA cloning
JVO-0895	5' P-ACATTTACCCTGTCTGTCC	sRNA cloning
JVO-0896	5' P-GAAAAAGGGTAGTCTCGC	sRNA cloning
JVO-0937	GTTTTTTTTAATACGACTCACTATAGGAAAGACGCGCATTT	T7 template

continued on next page

Name	Sequence	Used for
JVO-0938	AAAGGCCACTCACGG	T7 template
JVO-0941	GTTTTTTTTAATACGACTCACTATAGGACTTCCTGAGCCGG	T7 template
JVO-0942	AAAGCACCGCAATATG	T7 template
JVO-1034	GTTTTTTTTAATACGACTCACTATAGGATGAGGGGCATTTTATG	T7 template
JVO-1035	TTGCTGCAACGGTCAT	T7 template, toeprinting
JVO-1037	GTTTTTTTTAATACGACTCACTATAGGATCGACGAAAGGCGAT	T7 template
JVO-1038	GATGAGCGCAGTGAGTATT	T7 template
JVO-1039	GTTTTTTTTAATACGACTCACTATAGGATAACACTGCACGCGC	T7 template
JVO-1040	CCATCTTCTGCGTGC	T7 template
JVO-1048	GTTTTTTTTAATACGACTCACTATAGGATCAGCAGGACGCACT	T7 template
JVO-1049	TAACATCACCATCTAATTCAAC	T7 template
JVO-1060	GTTTTTTTTAATACGACTCACTATAGGAGGACAATATTGCAAC GTT	T7 template
JVO-1061	CAATACGAACCGTTTGC	T7 template
JVO-1062	GTTTTTTTTAATACGACTCACTATAGGCATTAATGAGTCAGTA AAAAGC	T7 template
JVO-1063	GTTTTTTTTAATACGACTCACTATAGGATCTATAGCGAAAAGC AGAATA	T7 template
JVO-1064	CCTGCGAGACTGCTAAT	T7 template
JVO-1065	GTTTTTTTTAATACGACTCACTATAGGAGTATGCTGCTAAAGC AC	T7 template
JVO-1066	GAATGCCATATGGCTTAAT	T7 template
JVO-1067	GTTTTTTTTAATACGACTCACTATAGGATCAGAATAGCACCCT GC	T7 template
JVO-1068	TGCCAAAATGTAATGTTCTG	T7 template
JVO-1117	TCAGCCATTTTGTGCGCTT	qRT-PCR
JVO-1118	TTCAGGATCGACAACGCCTT	qRT-PCR
JVO-1254	CCGACAAGCAAACGTTGGTAC	qRT-PCR
JVO-1255	TCACGGCTGACGTTCGATT	qRT-PCR
JVO-1256	TGCCGGATCTGATTAGCGA	qRT-PCR
JVO-1257	TGGCTAAATCGGCAAGGAAC	qRT-PCR
JVO-1271	GTTTTATGCATATCTATAGCGAAAAGCAGAATA	GFP fusion cloning
JVO-1340	CGCCTGGTAGATATCGAGCAA	qRT-PCR
JVO-1341	AATACGGCGCAGTGCGTTA	qRT-PCR
JVO-1381	GGACGCGACTGCTGACTAAAA	qRT-PCR
JVO-1382	AGAATTTCAGAGGTCGTCCCG	qRT-PCR
JVO-1628	CGGAAATCGCCAAATACCTG	qRT-PCR
JVO-1629	CACGCCGAACTCAAATCCTT	qRT-PCR
JVO-1775	ATCTTCTGCGTGCGCAA	toeprinting
JVO-1973	5’ P-TGCGCGTGCAGTGTTAT	GFP fusion cloning
JVO-1974	5’ P-AGCTATCAATGCGTCGACG	GFP fusion cloning
JVO-2154	5’ P-AACAACATCACAATACACGCTTACAAATTGTTGC	GFP fusion cloning
JVO-2155	ATTGCCTGCAACTATTCTTAAAAAAGCATGCATGT	GFP fusion cloning
JVO-2233	GTTTTTTTTAATACGACTCACTATAGGCTTTTTTAAGAATACA CGCTTACA	T7 template
JVO-2234	GTTTTTTTTAATACGACTCACTATAGGCTTTTTTAAGAATAGT TGCAGG	T7 template
JVO-2326	AGTACATTATGCGTACCGCCG	qRT-PCR
JVO-2327	AATGATAGCGATGCGCTGC	qRT-PCR
JVO-2328	TGGAATCGCTGAAAGGCAAG	qRT-PCR
JVO-2329	GCCACCACATCCACACCTTTA	qRT-PCR
JVO-2330	ATGCAAGCGGAGTGCTCATT	qRT-PCR
JVO-2331	TCATATCCATACCGGCGATCA	qRT-PCR

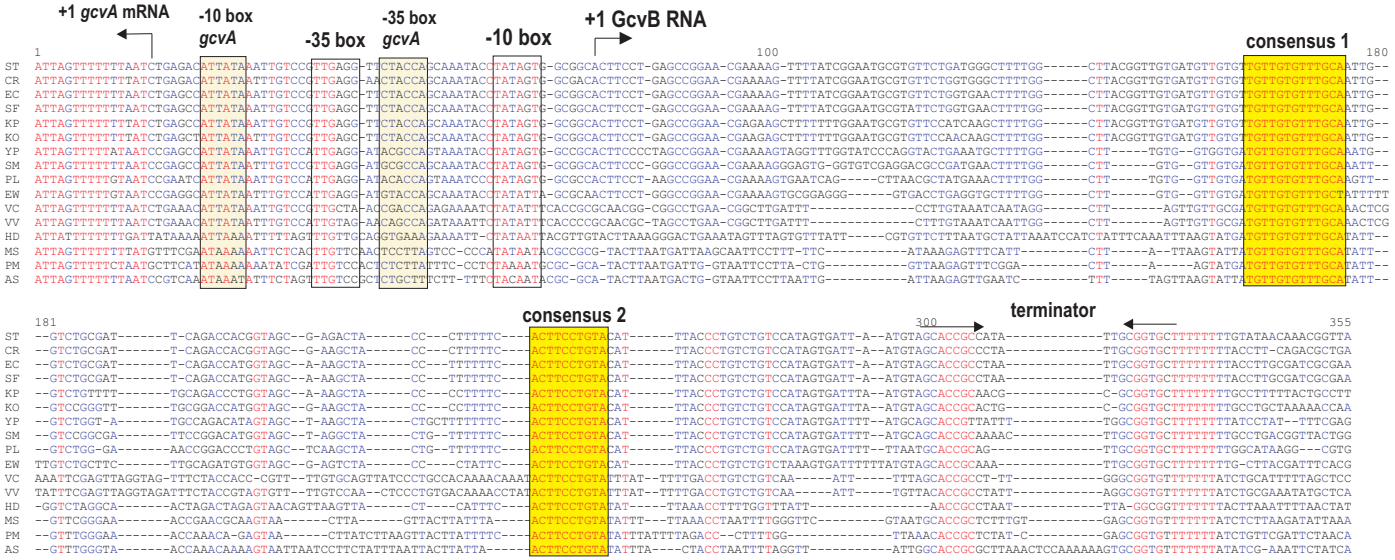


Figure 10.1: Multiple alignment of *gcvB* genes in different eubacteria. To identify *gcvB* homologues in diverse γ-proteobacteria, BLAST searches for the divergently encoded GcvA protein were carried out. The GcvA upstream regions corresponding to the genomic location of *gcvB* in *E. coli* and *Salmonella* were extracted, and transcriptional terminators of possible *gcvB* genes were predicted by RNAmotif as described in Chen *et al.* (2002). Based on sequence conservation and the assumption that the *gcvB* location is conserved, GcvB homologues could be identified in a wide range of eubacteria: **ST:** *Salmonella typhimurium*, **CR:** *Citrobacter rodentium*, **EC:** *Escherichia coli*, **SF:** *Shigella flexneri*, **KP:** *Klebsiella pneumoniae*, **KO:** *Klebsiella oxytoca*, **YP:** *Yersinia pestis*, **SM:** *Serratia marcescens*, **PL:** *Photorhabdus luminescens*, **EW:** *Erwinia carotovora*, **VC:** *Vibrio cholerae*, **VV:** *Vibrio vulnificus*, **HD:** *Haemophilus ducreyi*, **MS:** *Mannheimia succiniciproducens*, **PM:** *Pasteurella multocida*, **AS:** *Actinobacillus succinogenes*. Promoter elements of *gcvA* and *gcvB* genes are marked by black boxes; the transcriptional start sites are indicated by arrows. The stem-loop of the *gcvB* transcriptional terminator is indicated by two arrows. The two highly conserved regions, consensus R1 and R2, that are present in all GcvB homologues are highlighted by yellow boxes.

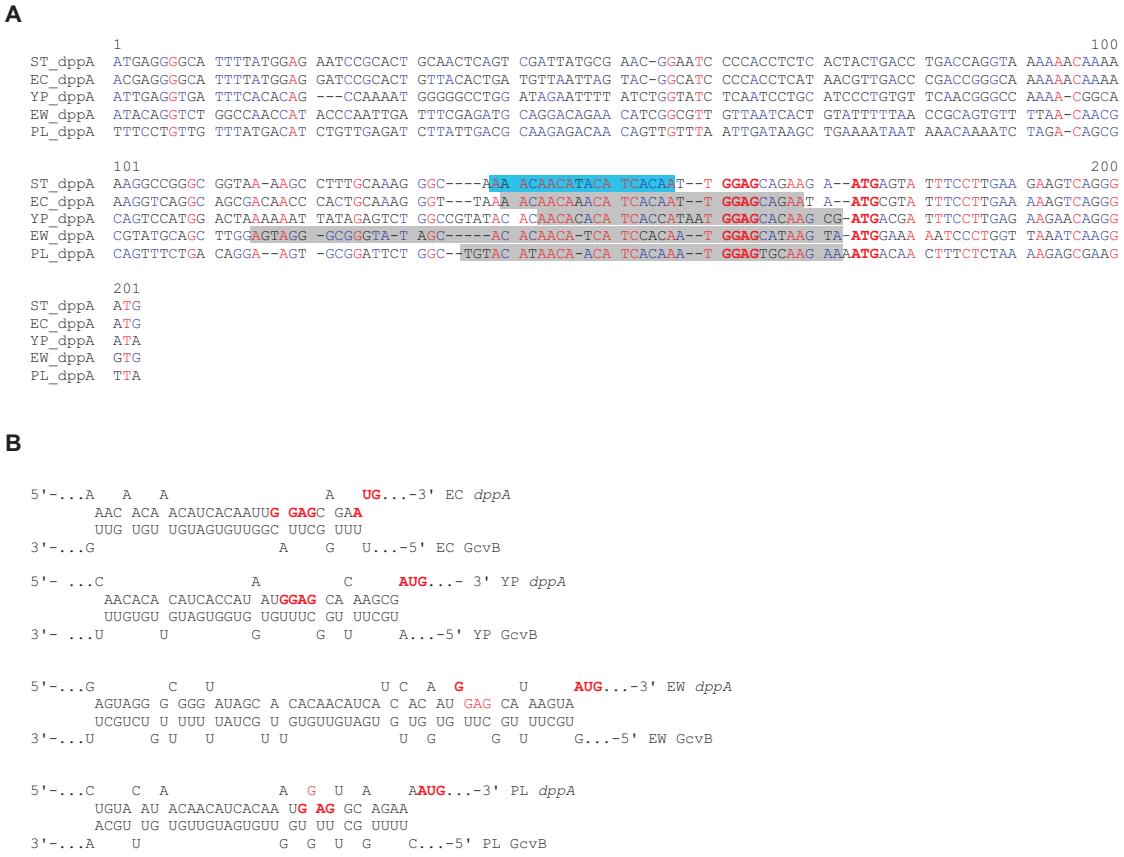


Figure 10.2: Alignment of *dppA* leaders. (A) Alignment of *dppA* leaders of diverse enterobacteria indicates conservation of the GcvB interaction site. **ST:** *Salmonella typhimurium*, **EC:** *Escherichia coli*, **YP:** *Yersinia pestis*, **EW:** *Erwinia carotovora*, **PL: Photorhabdus luminescens**. Shine-Dalgarno and start codon (ATG) are indicated in red bold letters. **(B)** Mapped GcvB interaction site for *Salmonella* and predicted interaction sites are highlighted in blue and grey, respectively.

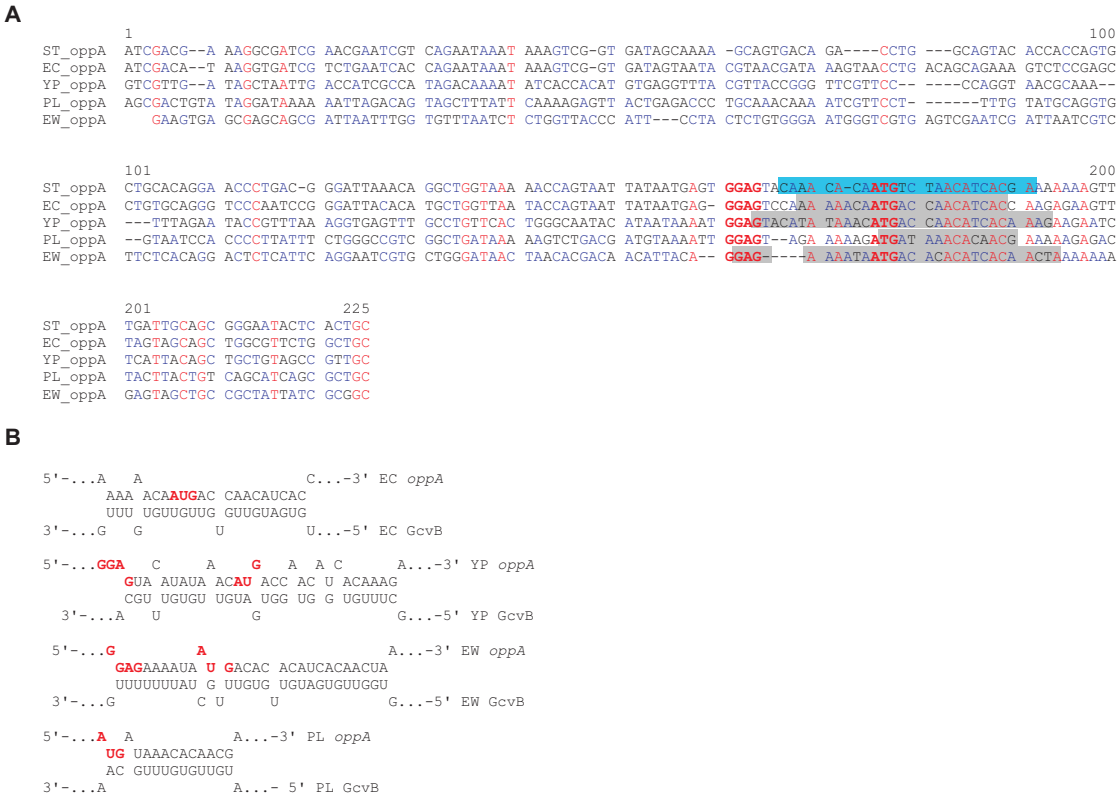


Figure 10.3: Alignment of *oppA* leaders. (A) Alignment of *oppA* leaders of diverse enterobacteria indicates conservation of the GcvB interaction site. **ST:** *Salmonella typhimurium*, **EC:** *Escherichia coli*, **YP:** *Yersinia pestis*, **EW:** *Erwinia carotovora*, **PL: Photorhabdus luminescens**. Shine-Dalgarno and start codon (ATG) are indicated in red bold letters. **(B)** Mapped GcvB interaction site for *Salmonella* and predicted interaction sites are highlighted in blue and grey, respectively.

Figure 10.4: Alignment of *gltI* leaders. (A) Alignment of *gltI* leaders of diverse enterobacteria indicates conservation of the GcvB interaction site. **ST:** *Salmonella typhimurium*, **EC:** *Escherichia coli*, **YP:** *Yersinia pestis*, **EW:** *Erwinia carotovora*, **PL: Photorhabdus luminescens**. Shine-Dalgarno and start codon (ATG) are indicated in red bold letters. **(B)** Mapped GcvB interaction site for *Salmonella* and predicted interaction sites are highlighted in blue and grey, respectively.

Figure 10.5: Alignment of *livK* leaders. (A) Alignment of *livK* leaders of diverse enterobacteria indicates conservation of the GcvB interaction site. **ST:** *Salmonella typhimurium*, **EC:** *Escherichia coli*, **YP:** *Yersinia pestis*, **EW:** *Erwinia carotovora*, **PL: Photorhabdus luminescens**. Shine-Dalgarno and start codon (ATG) are indicated in red bold letters. **(B)** Mapped GcvB interaction site for *Salmonella* and predicted interaction sites are highlighted in blue and grey, respectively.

Figure 10.6: Alignment of *livJ* and *argT* leaders. (A, B) Alignment of *livJ* and *argT* leaders of *Salmonella* and *E. coli* indicates conservation of the GcvB interaction site. **ST:** *Salmonella typhimurium*, **EC:** *Escherichia coli*. Shine-Dalgarno and start codon (ATG) are indicated in red bold letters. Mapped GcvB interaction site for *Salmonella* and predicted interaction sites are highlighted in blue and grey, respectively.

10.2. Appendix to Chapter 4

This section contains a supplementary Table to Chapter 4.

Table 10.4: DNA oligonucleotides used in this study.

Sequences are given in 5' → 3' direction and 5'P denotes a 5'-mono-phosphate.

Name	Sequence	Used for
JVO-0367	ACTGACATGGAGGAGGGA	GFP fusion cloning
JVO-1275	GTTTTGCTAGCTAGCGACTGTTCAGCCG	GFP fusion cloning
JVO-2850	GTTTTTATGCATGCCGTTTCCCCTCCAAT	GFP fusion cloning
JVO-2851	GTTTTTGCTAGCGTTCAAACTGACGCTTTCAGTT	GFP fusion cloning
JVO-2971	GTTTTTATGCATGCAAACACTTTGTTACATCCTG	GFP fusion cloning
JVO-3087	GTTTTTGCTAGCTGATGCTCGTTGCGTAGC	GFP fusion cloning
JVO-2058	GTTTTTATGCATTAACGTTAAACACAACACAAAT	GFP fusion cloning
JVO-2059	GTTTTTGCTAGCACCCACTGCCGCA	GFP fusion cloning
JVO-2806	GTTTTATGCATGCAAATACATATTCTGATAAAACG	GFP fusion cloning
JVO-2807	GTTTTTGCTAGCGAGGAACGATTCCAGAGAAC	GFP fusion cloning
JVO-2969	GTTTTTATGCATTTAATTTCACTTGCGACTTTG	GFP fusion cloning
JVO-2970	GTTTTTGCTAGCCCAGCCGATAAAACCAAC	GFP fusion cloning
JVO-2800	GTTTTATGCATGGAAGAAAAAACTGTGTTATGTATGT	GFP fusion cloning
JVO-2801	GTTTTTGCTAGCGATACGGTCGAGATCTTTGC	GFP fusion cloning
JVO-2804	GTTTTATGCATATTCGCACAGATAGCAATCTG	GFP fusion cloning
JVO-2805	GTTTTTGCTAGCGCGCAGATTCAGTGTATTAAAG	GFP fusion cloning
JVO-2973	GTTTTTATGCATAAAAAATAACAGGAGCATGACAC	GFP fusion cloning
JVO-2874	GTTTTTGCTAGCCAGCGCCTGTAGTGTTCTG	GFP fusion cloning
JVO-2842	GTTTTTATGCATTCAGGTGCTGTCATTACGAC	GFP fusion cloning
JVO-2843	GTTTTTGCTAGCGATATCGCGCGATTTTAACT	GFP fusion cloning
JVO-3388	GTTTTTATGCATAGTTAAGTAAACTGGTAGATGTTGC	GFP fusion cloning
JVO-3389	GTTTTTGCTAGCATTGAACCAAATATAATCAGCTTT	GFP fusion cloning
JVO-3378	GTTTTTATGCATATACAAGACAGACAAATAAAAATGAC	GFP fusion cloning
JVO-3379	GTTTTTGCTAGCACCGTTACCTGTGGTAATGG	GFP fusion cloning
JVO-3380	GTTTTTATGCATATTTGGCAATCAAGACGTT	GFP fusion cloning
JVO-3381	GTTTTTGCTAGCGGTAACGACGCGGATCT	GFP fusion cloning
JVO-2965	GTTTTTATGCATCTGACATAACAACAGAACATATTTCA	GFP fusion cloning
JVO-2808	GTTTTTGCTAGCGTTGGGTTTAATGATGGAAAAAG	GFP fusion cloning
JVO-2967	GTTTTTATGCATTTCATCTGCTATTTCCCGC	GFP fusion cloning
JVO-2968	GTTTTTGCTAGCTTTATCTTTCTCCAGCGATACCT	GFP fusion cloning
JVO-3330	TTTTTGCTAGCGTCGGCTGCGACTTTTA	GFP fusion cloning
JVO-0323	CCCTTTGCTAGCAAAGGAGAAGAACTTTTCACTG	GFP fusion cloning
JVO-0749	TTCGTTCCGGCTCAGGA	Northern blot probe
JVO-0750	AATCACTATGGACAGACAGGGTA	Northern blot probe
JVO-0322	CTACGGCGTTTCACTTCTGAGTTC	Northern blot probe
JVO-0897	5'P-ACTTCCTGAGCCGGAAC	sRNA cloning
PZE-XbaI	TCGTTTTATTTGATGCCTCTAGA	sRNA cloning
PLlaco-D	GTGCTCAGTATCTTGTTATCCG	sRNA cloning
JVO-0895	5'P-ACATTTACCCTGTCTGTCC	sRNA cloning
JVO-0896	5'P-GAAAAAGGGTAGTCTCGC	sRNA cloning
JVO-2856	5'P-TGCAAACACAACAACACAA	sRNA cloning
JVO-2857	GGTAGCGAGACTACCCTTTT	sRNA cloning
JVO-2858	CACTTCCTGTACATTTACCCTG	sRNA cloning

continued on next page

Name	Sequence	Used for
JVO-0745	5'P-ACCGTAAGCCAAAAGC	sRNA cloning
JVO-2989	5'P-CACTTCCTGTACATTTACCCTG	sRNA cloning
JVO-0746	5'P-CAATTGGTCTGCGATTC	sRNA cloning
JVO-2986	AGCACCGCCATATTGC	sRNA cloning
JVO-3327	ACTTCCTCTACATTTACCCTGTGTGTCCATAGTGATTAATGTAG CAC	sRNA cloning
JVO-3328	ATGGACACACAGGGTAAATGTAGAGGAAGTGGTGCTCAGTAT CT	sRNA cloning
JVO-2990	5'P-GGAACGAAAAGTTTTATCGG	sRNA cloning
JVO-0743	5'P-GCCGCCACTATAGGTATT	sRNA cloning
JVO-3355	TTGTTATCCGCTCACAATTC	sRNA cloning
JVO-1396	5'P-ACTTCCTGAGCCGGAAC	sRNA cloning
JVO-3466	AGCACACCTCCTGAGCCGGAACG	sRNA cloning
JVO-3467	CTCAGGAGGTGTGCTCAGTATCTTGTTATCC	sRNA cloning
JVO-3464	ACTTCCTCAGCCGGAACGAAAAGTTT	sRNA cloning
JVO-3465	TCCGGCTGAGGAAGTGTGCTCAGTATCTTGT	sRNA cloning
JVO-3468	TCCTGAGGCGGAACGAAAAGTTTTATCG	sRNA cloning
JVO-3469	CGTTCCGCCTCAGGAAGTGTGCTCAGTATCT	sRNA cloning
JVO-0941	GTTTTTTTTAATACGACTCACTATAGGACTTCCTGAGCCGG	T7 template
JVO-0942	AAAGCACCGCAATATG	T7 template
JVO-0742	5'P-AAAAAAGGGTAGTCTCGCTAC	T7 template
JVO-0937	GTTTTTTTTAATACGACTCACTATAGGAAAGACGCGCATTT	T7 template
JVO-0938	AAAGGCCACTCACGG	T7 template
JVO-1274	GTTTTTTTTAATACGACTCACTATAGGTCCTGATAACAGGATC GT	T7 template
JVO-1042	TATGACGGTTTGTAAGATTG	T7 template
JVO-1976	TCACCATCTAATTCAACAAGAATTG	T7 template, toeprinting

10.3. Appendix to Chapter 5

This section contains supplementary Tables to Chapter 5.

Table 10.5: *Salmonella* rRNAs, tRNAa, and housekeeping RNAS listed in the annotation file `LT2_rRNA_tRNA_hkRNAs.txt`.

Product Name	Start	End	Strand	Length	GeneID	Locus	Locus_tag	Links
16S ribosomal RNA	289189	290732	+	1544	1251767	rrsH	STM0249	-
Ile tRNA	290800	290873	+	74	1251768	ileV	STM0250	-
Asx tRNA	290986	291058	+	73	1251769	alaV	STM0251	-
23S ribosomal RNA	291244	294336	+	3093	1251770	rrlH	STM0252	-
5S ribosomal RNA	294519	294640	+	122	1251771	rrfH	STM0253	-
Val tRNA	294838	294911	+	74	1251772	aspU	STM0254	-
Sec tRNA	304277	304350	+	74	1251783	aspV	STM0265	-
Thr tRNA	368806	368878	+	73	1251842	thrW	STM0323	-
4.5S ribosomal RNA	524424	524539	+	116	1251987	ffs_SRP	STM0467_SRP	-
Arg tRNA	613567	613640	+	74	1252073	argU	STM0553	-
-	617984	618056	+	73	2673761	-	STM05559.T1	-
Gln tRNA	738011	738082	-	72	1252193	glnX	STM0673	-
Gln tRNA	738129	738200	-	72	1252194	glnV	STM0674	-
Xaa tRNA	738250	738323	-	74	1252195	metU	STM0675	-
Gln tRNA	738342	738413	-	72	1252196	glnW	STM0676	-
Gln tRNA	738452	738523	-	72	1252197	glnU	STM0677	-
Leu tRNA	738550	738631	-	82	1252198	leuW	STM0678	-
Met tRNA	738643	738716	-	74	1252199	metT	STM0679	-
Lys tRNA	818775	818847	+	73	1252271	lysT	STM0751	-
Val tRNA	818982	819054	+	73	1252272	valT	STM0752	-
Lys tRNA	819061	819133	+	73	1252273	lysW	STM0753	-
Lys tRNA	819186	819258	+	73	1252274	lysY	STM0754	-
Lys tRNA	819396	819468	+	73	1252275	lysZ	STM0755	-
Ser tRNA	1027441	1027525	-	85	1252468	serW	STM0949	-
Ser tRNA	1175321	1175405	+	85	1252604	serT	STM1086	-
Ser tRNA	1224743	1224827	-	85	1252652	serX	STM1134	-
Xaa tRNA	1333209	1333282	+	74	1252765	-	STM1247	-
Xaa tRNA	1345647	1345720	+	74	1252780	-	STM1262	-
Val tRNA	1501640	1501713	-	74	1252941	valW	STM1423	-
Val tRNA	1501727	1501800	-	74	1252942	valV	STM1424	-
Tyr tRNA	1852570	1852651	+	82	1253276	tyrT	STM1757	-
Tyr tRNA	1852859	1852940	+	82	1253278	tyrV	STM1759	-
Leu tRNA	2035366	2035449	-	84	1253463	leuZ	STM1942	-
Cys tRNA	2035464	2035534	-	71	1253464	cysT	STM1943	-
Gly tRNA	2035590	2035662	-	73	1253465	glyW	STM1944	-
Ser tRNA	2082177	2082263	-	87	1253521	serU	STM2000	-
Asn tRNA	2083250	2083322	+	73	1253523	asnT	STM2002	-
Asn tRNA	2084220	2084292	+	73	1253525	asnT	STM2004	-
Asn tRNA	2094165	2094237	-	73	1253533	asnW	STM2012	-
Asn tRNA	2095995	2096067	+	73	1253535	asnU	STM2014	-
Pro tRNA	2330749	2330822	+	74	1253751	tRNAPro2	STM2229	-
Arg tRNA	2505822	2505893	+	72	1253916	argW	STM2394	-
Ala tRNA	2528939	2529011	-	73	1253933	alaX	STM2411	-
Ala tRNA	2529057	2529129	-	73	1253934	alaW	STM2412	-
Val tRNA	2531835	2531907	+	73	1253938	valU	STM2416	-
Val tRNA	2531956	2532028	+	73	1253939	valX	STM2417	-
Val tRNA	2532073	2532145	+	73	1253940	valY	STM2418	-
Lys tRNA	2532153	2532225	+	73	1253941	lysV	STM2419	-
Xaa tRNA	2761647	2761720	-	74	1254138	-	STM2615	
5S ribosomal RNA	2796440	2796561	-	122	1254179	rrfG	STM2656	-
23S ribosomal RNA	2796755	2799764	-	3010	1254180	rrlG	STM2657	-
Sec tRNA	2799958	2800030	-	73	1254181	gltW	STM2658	-

continued on next page

Product Name	Start	End	Strand	Length	GeneID	Locus	Locus_tag	Links
16S ribosomal RNA	2800118	2801660	-	1543	1254182	rrsG	STM2659	-
regulatory RNA	2843947	2844309	+	363	1254216	tmRNA	STM2693_tmRNA	-
Arg tRNA	2969928	2970001	-	74	1254344	-	STM2821	-
Arg tRNA	2970198	2970271	-	74	1254345	-	STM2822	-
Arg tRNA	2970337	2970410	-	74	1254346	-	STM2823	-
Arg tRNA	2970476	2970549	-	74	1254347	-	STM2824	-
Ser tRNA	2970556	2970645	-	90	1254348	-	STM2825	-
Met tRNA	3141064	3141137	+	74	1254512	metZ	STM2989	-
Met tRNA	3141170	3141243	+	74	1254513	metW	STM2990	-
Gly tRNA	3197425	3197495	-	71	1254560	glyU	STM3037	-
Phe tRNA	3276146	3276218	+	73	1254639	pheV	STM3116	-
Ile tRNA	3379004	3379076	+	73	1254736	ileX	STM3213	-
regulatory RNA	3414805	3415177	-	377	1254769	rnpB	STM3246_RNaseP	-
Met tRNA	3458506	3458579	-	74	1254812	metY	STM3289	-
Leu tRNA	3461755	3461838	-	84	1254815	leuU	STM3292	-
5S ribosomal RNA	3566622	3566743	-	122	1254916	rrfF	STM3393	-
Thr tRNA	3566784	3566856	-	73	1254917	thrV	STM3394	-
5S ribosomal RNA	3566867	3566988	-	122	1254918	rrfD	STM3395	-
23S ribosomal RNA	3567078	3570071	-	2994	1254919	rrlD	STM3396	-
Gln tRNA	3570307	3570379	-	73	1254920	-	STM3397	-
16S ribosomal RNA	3570463	3572006	-	1544	1254921	rrsD	STM3398	-
Pro tRNA	3820412	3820485	-	74	1255158	proK	STM3634	-
Sec tRNA	3948576	3948666	+	91	1255275	selC	STM3751	-
16S ribosomal RNA	4100132	4101675	+	1544	1255416	rrsC	STM3889	-
Trp tRNA	4101759	4101831	+	73	1255417	gltU	STM3890	-
23S ribosomal RNA	4102028	4105021	+	2994	1255418	rrlC	STM3891	-
5S ribosomal RNA	4105111	4105230	+	120	1255420	rrfC	STM3894	-
Sec tRNA	4105283	4105362	+	80	1255421	aspT	STM3895	-
Xaa tRNA	4105368	4105440	+	73	1255422	trpT	STM3896	-
Arg tRNA	4140908	4140981	+	74	1255457	argX	STM3931	-
His tRNA	4141039	4141111	+	73	1255458	hisR	STM3932	-
Leu tRNA	4141135	4141218	+	84	1255459	leuT	STM3933	-
Pro tRNA	4141264	4141337	+	74	1255460	proM	STM3934	-
16S ribosomal RNA	4196059	4197600	+	1542	1255514	rrsA	STM3988	-
Ile tRNA	4197670	4197743	+	74	1255515	ileT	STM3989	-
Asx tRNA	4197858	4197930	+	73	1255516	alaT	STM3990	-
23S ribosomal RNA	4198116	4201110	+	2995	1255517	rrlA	STM3991	-
5S ribosomal RNA	4201302	4201423	+	122	1255518	rrfA	STM3992	-
16S ribosomal RNA	4351130	4352673	+	1544	1255658	rrsB	STM4132	-
Val tRNA	4352741	4352814	+	74	1255659	ileU	STM4133	-
Ala tRNA	4352929	4353001	+	73	1255660	-	STM4134	-
23S ribosomal RNA	4353187	4356180	+	2994	1255661	rrlB	STM4135	-
5S ribosomal RNA	4356373	4356493	+	121	1255662	rrfB	STM4136	-
Thr tRNA	4360046	4360118	+	73	1255668	thrU	STM4142	-
Tyr tRNA	4360130	4360211	+	82	1255669	tyrU	STM4143	-
Gly tRNA	4360331	4360402	+	72	1255670	glyT	STM4144	-
Thr tRNA	4360412	4360484	+	73	1255671	thrT	STM4145	-
16S ribosomal RNA	4394675	4396219	+	1545	1255703	rrsE	STM4177	-
Sec tRNA	4396303	4396375	+	73	1255704	gltV	STM4178	-
23S ribosomal RNA	4396611	4399604	+	2994	1255705	rrlE	STM4179	-
5S ribosomal RNA	4399797	4399918	+	122	1255706	rrfE	STM4180	-
Phe tRNA	4566577	4566649	-	73	1255847	pheR	STM4321	-
Gly tRNA	4596412	4596484	+	73	1255878	glyV	STM4352	-
Gly tRNA	4596644	4596716	+	73	1255879	glyX	STM4353	-
Gly tRNA	4596876	4596948	+	73	1255880	glyY	STM4354	-
Leu tRNA	4730996	4731077	+	82	1256013	leuX	STM4487	-
Leu tRNA	4810613	4810696	-	84	1256079	leuV	STM4553	-
Leu tRNA	4810731	4810814	-	84	1256080	leuP	STM4554	-
Leu tRNA	4810845	4810928	-	84	1256081	leuQ	STM4555	-

Table 10.6: Known *Salmonella* sRNAs listed in the annotation file `LT2_known_sRNAs.txt`.

Product Name	Start	End	Strand	Length	GeneID	Locus	Locus_tag	Links
sRNA	13507	13591	+	84	1	IGR_dnaK_dnaJ	tpke11	-
sRNA	126707	126796	-	90	2	IGR_tbpA_yabN	sroA	-
sRNA	128574	128812	+	239	3	IGR_yabN_leuD	sgrS	-
sRNA	176082	176238	-	157	4	IGR_pdhR_aceE	tp2	-
sRNA	254130	254268	+	139	5	IGR_map_rpsB	t44	-
sRNA	505384	505439	-	56	6	IGR_clpX_lon	sraA	-
sRNA	556005	556085	+	81	8	IGR_ybaK_ybaP	sroB	-
sRNA	728761	728913	-	153	9	IGR_gltJ_gltI	sroC	-
sRNA	902040	902128	-	89	10	IGR_ybiP_STM0835	rybA	-
sRNA	942554	942632	-	79	11	IGR_STM0869_STM0870	rybB	-
sRNA	1275071	1275236	+	166	12	IGR_yceF_yceD	sraB	-
sRNA	1444832	1444938	-	107	13	IGR_ydiL_ydiK	rprA	-
sRNA	1450415	1450519	+	105	14	IGR_ydiH_STM1368	rydB	-
sRNA	1729673	1729738	+	66	15	IGR_STM1638_cybB	rydC	-
sRNA	1745678	1745786	-	109	16	IGR_nifJ_ynaF	micC	-
sRNA	1968053	1968155	-	103	17	IGR_STM1871_STM1872	ryeB	-
sRNA	2068649	2068736	-	88	18	IGR_yodD_yedP	dsrA	-
sRNA	2077175	2077269	+	95	19	IGR_STM1994_ompS	rseX	-
sRNA	2213871	2214016	+	146	21	IGR_yegD_STM2126	ryeC	-
sRNA	2231130	2231216	+	87	20	IGR_yegQ_STM2137	ryeE	-
sRNA	2366913	2367005	+	93	22	IGR_ompC_yojN	micF	-
sRNA	2515608	2516006	-	399	23	IGR_ddg_yfdZ	tpke70	-
sRNA	2674934	2675228	+	295	24	IGR_STM2564_sseB	ryfA	-
sRNA	2707664	2707847	-	183	25	IGR_yfhK_purG	sroF	-
sRNA	2966853	2966926	+	74	27	IGR_luxS_gshA	micA	-
sRNA	2987638	2987745	+	108	28	IGR_ygbD_hypF	C0664	-
sRNA	3044924	3045015	+	93	47	IGR_invH_STM2901	invR	-
sRNA	3116697	3117059	-	363	29	IGR_yqcC_syd	csrB	-
sRNA	3135317	3135522	+	206	30	IGR_gcvA_ygdL	gcvB	-
sRNA	3170122	3170208	-	87	31	IGR_aas_galR	omrA	-
sRNA	3170323	3170408	-	86	32	IGR_aas_galR	omrB	-
sRNA	3222098	3222280	+	183	33	IGR_ygfE_ygfA	ssrS	-
sRNA	3222913	3223065	+	153	34	IGR_ygfA_serA	rygC	-
sRNA	3362327	3362474	-	148	35	IGR_yqiK_rfaE	rygD	-
sRNA	3392069	3392261	+	193	36	IGR_ygjR_ygjT	sraF	-
sRNA	3451437	3451607	+	174	38	IGR_pnp_rpsO	sraG	-
sRNA	3490383	3490500	+	118	39	IGR_yhbL_arcB	sraH	-
sRNA	3715401	3715495	-	95	40	IGR_yhhX_yhhY	ryhB	-
sRNA	3998018	3998147	-	130	41	IGR_ilvB_emrD	istR	-
sRNA	4141650	4141854	+	205	42	IGR_yifK_hemY	sraJ	-
sRNA	4209066	4209175	+	110	43	IGR_polA_yihA	spot42	-
sRNA	4210157	4210400	+	244	44	IGR_yihA_yihI	csrC	-
sRNA	4342866	4342986	-	121	45	IGR_argH_oxyR	oxyS	-
sRNA	4504870	4505010	-	141	46	IGR_soxR_STM4267	sraL	-

Table 10.7: Predicted *Salmonella* sRNAs listed in the annotation file `sRNAs LT2_predicted_sRNAs.txt`.

Product Name	Start	End	Strand	Length	GeneID	Locus	Locus_tag	Links
sRNA	46050	46114	-	65	1	IGR_STM0038_nhaA	STnc10	-
sRNA	51926	52260	+	335	2	IGR_STM0042_rpsT	STnc20	-
sRNA	58792	58923	+	132	3	IGR_lytB_STM0050	STnc30	-
sRNA	161464	161537	+	74	4	IGR_secA_mutT	STnc40	-
sRNA	182458	182539	-	82	5	IGR_lpdA-STM0155	STnc50	-
sRNA	230063	230277	-	215	6	IGR_fhuB_stfA	STnc60	-
sRNA	670157	670305	+	149	7	IGR_dsbG_ahpC	STnc70	-
sRNA	967580	967900	+	321	8	IGR_STM0897_STM0898	STnc80	-
sRNA	974284	974363	+	80	9	IGR_STM0903_STM0904	STnc90	-
sRNA	975011	975224	+	214	10	IGR_STM0904_STM0905	STnc100	-
sRNA	976578	976765	+	188	11	IGR_STM0905_STM0906	STnc110	-
sRNA	1004432	1004777	-	346	12	IGR_STM0929_orfB	STnc120	-
sRNA	1045098	1045232	-	135	13	IGR_serS_dmsA	STnc130	-
sRNA	1113681	1113750	+	70	14	IGR_STM1025_STM1026	STnc140	-
sRNA	1325649	1325914	-	266	15	IGR_icdA_STM1239	STnc150	-
sRNA	1345732	1345782	-	51	16	IGR_STM1262_STM1263	STnc160	-
sRNA	1605784	1606116	-	333	17	IGR_STM1528_STM1530	STnc170	-
sRNA	1807565	1807776	-	212	18	IGR_acnA_cysB	STnc190	-
sRNA	1937518	1937652	+	135	19	IGR_STM1841_kdgR	STnc200	-
sRNA	1979550	1979598	-	49	20	IGR_edd_zwf	STnc210	-
sRNA	2032404	2032580	+	177	21	IGR_yecA_STM1939	STnc220	-
sRNA	2078990	2079068	-	79	22	IGR_ompS_cspB	STnc240	-
sRNA	2115370	2115452	+	83	23	IGR_pocR_pduF	STnc250	-
sRNA	2147333	2147409	-	77	24	IGR_yeeF_yeeY	STnc260	-
sRNA	2596789	2596882	-	94	25	IGR_acrD_yffB	STnc270	-
sRNA	2966073	2966247	+	175	26	IGR_STM2816_luxS	STnc290	-
sRNA	3179540	3179622	+	82	27	IGR_kduI_yqeF	STnc310	-
sRNA	3194914	3194996	-	83	28	IGR_tnpA_4_STM3033	STnc320	-
sRNA	3283807	3283965	-	159	29	IGR_STM3123_STM3124	STnc330	-
sRNA	3393267	3393327	-	61	30	IGR_ygjT_ygjU	STnc340	-
sRNA	3404895	3404949	+	55	31	IGR_yhaO_tdcG	STnc350	-
sRNA	3468497	3468553	-	57	32	IGR_greA_dacB	STnc360	-
sRNA	3635756	3635884	-	129	33	IGR_tnpA_5_yhfL	STnc370	-
sRNA	3761373	3761440	-	68	34	IGR_uspA_yhiP	STnc380	-
sRNA	3780254	3780402	+	149	35	IGRyhjB_yhjC	STnc390	-
sRNA	3839688	3839758	+	71	36	IGR_STM3654_glyS	STnc400	-
sRNA	3885629	3885736	-	108	37	IGR_STM3691_lldP	STnc410	-
sRNA	3902594	3902653	-	60	38	IGR_yibD_tdh	STnc420	-
sRNA	4051145	4051340	+	196	39	IGR_STM3844_STM3845	STnc430	-
sRNA	4072507	4072730	+	224	40	IGR_glmU_STM3863	STnc440	-
sRNA	4251480	4251539	-	60	41	IGR_yiiG_STM4041	STnc450	-
sRNA	4441898	4442059	-	162	42	IGR_pgi_yjbE	STnc460	-
sRNA	4559193	4559277	+	85	43	IGR_STM4310_tnpA_6	STnc470	-
sRNA	4645079	4645134	-	56	44	IGR_ytfL_msrA	STnc480	-
sRNA	4758187	4758332	-	146	45	IGR_STM4503_STM4504	STnc490	-

Table 10.8: Coverage of known and candidate *Salmonella* sRNA loci in pyrosequencing data.

sRNA[1]	Alternative IDs[2]	Reference[3]	Adjacent genes[4]	Orientation[5]	Start[6]	End[6]	Reads coIP-Ctr[7]	Reads Hfq-coIP[8]	Enrichment[9]	Northern blot[10]
STnc10	-	V	STM0038/*nhaA*	→←→	46114	46050	0	0		np
STnc20	-	V	STM0042/*rpsT*	←→←	51926	52260	1	2	2.0	np
STnc30	-	V	*lytB*/STM005	→→→	58792	58923	1	0		np
STnc470	-	IV	STM0081/STM0082	→←←	94548	94770	0	70	≥70.0	≈1250nt
sgrS	*ryaA*	I	*yabN/leuD*	←→←	128574	128812	3	61	20.3	
STnc40	-	V	*secA/mutT*	→→→	161464	161537	0	0		np
STnc50	-	V	*lpdA*/STM0155	→←→	182539	182458	0	0		np
STnc60	-	V	*fhuB/stfA*	→←→	230277	230063	0	0		np
isrA	-	II	STM0294.ln/STM0295	→→→	339338	339760	0	0		
sroB	*rybC*	I	*ybaK/ybaP*	←→←	556005	556085	27	1530	56.7	
STnc480	-	IV	*glxK/ylbA*	→←←	587848	587926	4	74	18.5	nd
STnc70	-	V	*dsbG/ahpC*	←→→	670157	670305	5	7	1.4	np
sroC	-	I	*gltJ/gltI*	←←←	728913	728761	26	898	34.5	
rybB	p25	III	STM0869/STM0870	→←←	942632	942554	3	103	34.3	
STnc80	-	V	STM0897/STM0898	←→←	967580	967900	0	0		np
STnc90	-	V	STM0903/STM0904	→→←	974284	974363	0	0		np
STnc100	-	V	STM0904/STM0905	←→→	975011	975224	0	0		np
STnc110	-	V	STM0905/STM0906	→→→	976578	976765	0	0		np
STnc120	-	V	STM0929/*orfB*	←←→	1004777	1004432	0	0		np
STnc490[11]	-	IV	*clpA/tnpA*_1	→←→	1024975	1025165	75	385	5.1	≈85nt
STnc130	-	V	*serS/dmsA*	→←→	1045232	1045098	0	0		nd
isrB-1	-	II	*sbcA*/STM1010	←→←	1104179	1104266	2	4	2.0	
STnc140	-	V	STM1025/STM1026	←→←	1113681	1113750	0	0		np
STnc500	-	IV	STM1127/STM1128	←←←	1216157	1216440	7	84	12.0	≈65nt
sraB	pke2	I	*yceF/yceD*	←→→	1275071	1275236	0	0		
STnc640	-	IV	*icdA*/STM1239	→→→	1325636	1326082	0	10	≥10.0	≈1500nt
STnc150	-	V	*icdA*/STM1239	→←→	1325914	1325649	0	1	≥1.0	≈90nt
isrC	-	II	*envF/msgA*	←→←	1329145	1329432	0	1	≥1.0	
STnc510	-	IV	STM1245/*pagC*	→→→	1331440	1332250	4	28	7.0	nd
STnc520	-	IV	STM1248/STM1249	→←←	1332809	1334044	12	100	8.3	≈80nt
STnc160	-	V	STM1262/STM1263	→←→	1345782	1345732	0	0		np
isrD	-	II	STM1261/STM1263	→←→	1345788	1345738	0	0		
ryhB-2	*isrE*	II	STM1273/*yeaQ*	→←→	1352987	1352875	0	0		
STnc530	-	IV	*yeaJ/yeaH*	→←→	1359779	1360418	2	15	7.5	nd

continued on next page

sRNA[1]	Alternative IDs[2]	Reference[3]	Adjacent genes[4]	Orientation[5]	Start[6]	End[6]	Reads coIP-Ctr[7]	Reads Hfq-coIP[8]	Enrichment[9]	Northern blot[10]
STnc540	-	IV	*himA/btuC*	→→→	1419369	1419570	7	23	3.3	≈85nt
rprA	IS083	I	*ydiK/ydiL*	←←←	1444938	1444832	37	286	7.7	
rydB	tpe7, IS082	I	*ydiH*/STM1368	→→←	1450415	1450519	4	10	2.5	
STnc550	-	IV	*purR/sodB*	←→←	1508946	1509412	6	10	1.7	nd
STnc570[12]	*yneM* ORF	IV	*ydeI/ydeE*	→←←	1593723	1594413	2	21	10.5	≈190nt
STnc560		IV	*ydeI/ydeE*	→→←	1593723	1594413	10	290	29.0	≈90nt
STnc170	-	V	STM1528/STM1530	←←→	1606116	1605784	0	0		np
isrF	-	II	STM1552/STM1554	→←←	1630160	1629871	1	0		
rydC	IS067	I	STM1638/*cybB*	→→←	1729673	1729738	5	245	49.0	
micC	IS063, tke8	III	*nifJ/ynaF*	→←→	1745786	1745678	0	15	≥15.0	
STnc580	-	IV	*dbpA*/STM1656	←←←	1749662	1750147	11	311	28.3	≈100nt
STnc180	-	V	*acnA/cysB*	←←←	1807776	1807565	1	5	5.0	≈2000nt
STnc190	-	V	STM1841/*kdgR*	→→←	1937518	1937652	1	12	12.0	≈500nt
ryeB	tpke79	I	STM1871/STM1872	→←←	1968155	1968053	24	653	27.2	
STnc200	-	V	*edd/zwf*	←←←	1979598	1979550	0	3	≥3.0	nd
STnc210	-	V	*yecA*/STM1939	←→←	2032404	2032580	0	0		np
dsrA	-	I	*yodD/yedP*	→←→	2068736	2068649	6	149	24.8	
rseX	-	I	STM1994/*ompS*	←→→	2077175	2077269	0	3	≥3.0	
STnc220	-	V	*ompS/cspB*	→←←	2079068	2078990	0	8	≥8.0	nd
STnc230	-	V	*pocR/pduF*	←→←	2115370	2115452	0	0		np
STnc240	-	V	*yeeF/yeeY*	←←←	2147409	2147333	0	1	≥1.0	np
ryeC	tp11	I	*yegD*/STM2126	→→→	2213871	2214016	42	72	1.7	
cyaR	*ryeE*	III	*yegQ*/STM2137	→→→	2231130	2231216	31	659	21.3	
isrG	-	II	STM2243/STM2244	←→→	2344732	2345013	0	0		
micF	-	III	*ompC/yojN*	←→→	2366913	2367005	0	11	≥11.0	
isrH-2	-	II	*glpC*/STM2287	→←→	2394582	2394303	0	0		
isrH-1	-	II	*glpC*/STM2287	→←→	2394753	2394303	0	0		
STnc250[12]	*ypfM* ORF	V	*acrD/yffB*	→←→	2596882	2596789	6	24	4.0	≈220nt
ryfA	tp1	I	STM2534/*sseB*	→→←	2674934	2675228	3	6	2.0	
glmY	tke1, *sroF*	I	*yfhK/purG*	←←←	2707847	2707664	20	92	4.6	
isrI	-	II	STM2614/STM2616	→←←	2761576	2761329	0	2	≥2.0	
isrJ	-	II	STM2614/STM2616	→←←	2762031	2761957	1	0		
isrK	-	II	STM2616/STM2617	←←←	2762867	2762791	0	0		
isrB-2	-	II	STM2631/*sbcA*	→←→	2770965	2770872	0	0		
isrL	-	II	*smpB*/STM2690	→←→	2839399	2839055	0	0		
isrM	-	II	STM2762/STM2763	←→→	2905050	2905378	0	0		

continued on next page

sRNA[1]	Alternative IDs[2]	Reference[3]	Adjacent genes[4]	Orientation[5]	Start[6]	End[6]	Reads coIP-Ctr[7]	Reads Hfq-coIP[8]	Enrichment[9]	Northern blot[10]
isrN	-	II	STM2764/STM2765	←→←	2906925	2907067	0	0		
STnc260	-	V	STM2816/*luxS*	←→←	2966073	2966247	0	0		np
micA	*sraD*	I	*luxS/gshA*	←→←	2966853	2966926	1	128	128.0	
STnc590	-	IV	*avrA/sprB*	←←←	3010807	3010966	3	27	9.0	nd
STnc600	-	IV	*hilD/hilA*	→→→	3018766	3019855	3	68	22.7	nd
invR	STnc270	III	*invH*/STM2901	→→→	3044924	3045014	113	3236	28.6	
csrB	-	III	*yqcC/syd*	←←←	3117059	3116697	69	67		
gcvB	IS145	III	*gcvA/ygdI*	←→←	3135317	3135522	12	402	33.5	
omrA	*rygB*	III	*aas/galR*	←←→	3170208	3170122	0	51	≥51.0	
omrB	t59, *rygA*, *sraE*	III	*aas/galR*	←←→	3170408	3170322	1	52	52.0	
STnc280	-	V	*kduI/yqeF*	←→←	3179540	3179622	0	1	≥1.0	np
STnc290	-	V	*tnpA_4*/STM3033	←←←	3194996	3194914	2	72	36.0	≈ 85nt
isrO	-	II	STM3038/STM3039	←→→	3198380	3198580	0	0		
ssrS	-	I	*ygfE/ygfA*	→→→	3222098	3222280	836	451		
rygC	t27	I	*ygfA/serA*	→→←	3222913	3223065	14	17	1.2	
STnc300	-	V	STM3123/STM3124	←←→	3283965	3283807	0	0		np
rygD	tp8, C0730	I	*yqiK/rfaE*	→←←	3362474	3362327	17	104	6.1	
sraF	tpk1, IS160	I	*ygjR/ygjT*	→→→	3392069	3392261	0	25	≥25.0	
STnc310	-	V	*ygjT/ygjU*	→←→	3393327	3393267	0	0		np
STnc320	-	V	*yhaO/tdcG*	←→←	3404895	3404949	0	1	≥1.0	np
STnc610	-	IV	*yhbC/metY*	←←←	3458296	3458578	1	19	19.0	≈ 1250nt
STnc330	-	V	*greA/dacB*	←←→	3468553	3468497	1	12	12.0	≈ 1500nt
sraH	*ryhA*	I	*yhbL/arcB*	←→←	3490383	3490500	55	2292	41.7	
STnc340	-	V	*tnpA_5/yhfL*	←←→	3635884	3635756	0	0		nd
ryhB-1	*sraI*, IS176	I	*yhhX/yhhY*	←←→	3715495	3715401	0	2	≥2.0	
STnc350	-	V	*uspA/yhiP*	→←→	3761440	3761373	0	0		nd
STnc360	-	V	*yhjB/yhjC*	←→→	3780254	3780402	0	0		np
STnc370	-	V	STM3654/*glyS*	←→←	3839688	3839758	0	0		np
STnc380	-	V	STM3691/*lldP*	→←→	3885736	3885629	0	0		np
STnc390	-	V	*yibD/tdh*	←←←	3902653	3902594	0	0		nd
istR-1		VI	*ilvB/emrD*	←←→	3998147	3998018	0	0		≈ 75nt
istR-2		VI	*ilvB/emrD*	←←→	3998147	3998018	0	0		≈ 140nt
STnc400	-	V	STM3844/STM3845	→→→	4051145	4051340	112	42		≈ 55nt
STnc410	-	V	*glmU*/STM3863	←→←	4072507	4072730	0	0		np
glmZ	k19, *ryiA*, *sraJ*	I	*yifK/hemY*	→→←	4141650	4141854	20	196	9.8	
spf	*spf*	I	*polA/yihA*	→→←	4209066	4209175	2	33	16.5	

continued on next page

sRNA[1]	Alternative IDs[2]	Reference[3]	Adjacent genes[4]	Orientation[5]	Start[6]	End[6]	Reads coIP-Ctr[7]	Reads Hfq-coIP[8]	Enrichment[9]	Northern blot[10]
csrC	*sraK, ryiB*, tpk2	III	*yihA/yihI*	←→→	4210157	4210400	63	64		
STnc420	-	V	*yiiG*/STM4041	→←←	4251539	4251480	0	0		np
isrP	-	II	STM4097/STM4098	←→←	4306719	4306866	0	2	≥2.0	
oxyS	-	I	*argH/oxyR*	→←→	4342986	4342866	0	10	≥10.0	
STnc430	-	V	*pgi/yjbE*	→←→	4442059	4441898	0	0		np
STnc620	-	IV	*ssb*/STM4257	→→→	4476817	4477856	4	41	10.3	nd
sraL	*ryjA*	III	*soxR*/STM4267	→←→	4505010	4504870	0	0		
STnc630	-	IV	*proP/basS*	→→←	4532473	4532638	1	27	27.0	nd
STnc440	-	V	STM4310/*tnpA_6*	→→→	4559193	4559277	9	456	50.7	≈ 85nt
STnc450	-	V	*ytfL/msrA*	←←←	4645134	4645079	0	0		np
STnc460	-	V	STM4503/STM4504	→←→	4758332	4758187	0	0		np
isrQ	-	II	STM4508/STM4509	←→→	4762997	4763158	0	0		

[1] Gene names of *Salmonella* sRNAs that were identified in this and previous studies. The identification method is given in the third column. sRNA names follow the *Salmonella* and/or *E. coli* nomenclature referenced in Hershberg *et al.* (2003), Padalon-Brauch *et al.* (2008), and Papenfort *et al.* (2008).

[2] Alternative sRNA IDs. References in Hershberg *et al.* (2003), Padalon-Brauch *et al.* (2008), and Papenfort *et al.* (2008) except STnc490, 500, 520, 540, 560, 570, 580, which have been newly predicted in this study.

[3] Evidence for sRNAs in *Salmonella*. **(I)** Conserved sRNAs found in *Salmonella* cDNA libraries and previously shown to be expressed in *E. coli* (relevant ref. in Papenfort *et al.*, 2008; Table 1). **(II)** sRNA previously predicted and validated on Northern blots in *Salmonella* by Padalon-Brauch *et al.* (2008). **(III)** sRNA previously validated on Northern blots in *Salmonella* (Altier *et al.*, 2000a; Figueroa-Bossi *et al.*, 2006; Fortune *et al.*, 2006; Papenfort *et al.*, 2006, 2008; Pfeiffer *et al.*, 2007; Sharma *et al.*, 2007; Viegas *et al.*, 2007). **(IV)** sRNA predicted through cDNA sequencing and validated by Northern blot analysis in this study. **(V)** sRNA previously predicted by Pfeiffer *et al.* (2007) is recovered in cDNA sequences and validated by Northern blot analysis in this study. **(VI)** IstR sRNAs (Vogel *et al.*, 2004) were not recovered in cDNA sequences but their expression in *Salmonella* was validated by Northern blot analysis in the complete study (Sittka *et al.*, 2008).

[4] Flanking genes of the intergenic region in which the sRNA candidate is located.

[5] Orientation of sRNA candidate (middle) and flanking genes (→ and ← denote location of a gene on the clockwise or the counterclockwise strand of the *Salmonella* chromosome).

[6] Genomic location of sRNA candidate gene according to the *Salmonella typhimurium* LT2 genome. For STnc470 through STnc640 start and end of the entire intergenic region are given.

[7] Out of 145,873 sequences in total.

[8] Out of 122,326 sequences in total.

[9] Enrichment factor calculated by dividing the number of reads from Hfq coIP by the number of reads from the control coIP.

[10] Denotes verification on Northern blot in this study for new RNA transcripts; the estimated size is given in nucleotides (np = not probed; nd = no detectable transcript).

[11] The cDNA reads map antisense internally of the IS200 element. Based on sequence identity they map to all IS200 elements (*tnpA_1* to *tnpA_6*).

[12] STnc250 and STnc570 contain small ORFs annotated as *ypfM* and *yneM*, respectively, in *E. coli* (Wassarman *et al.*, 2001).

Table 10.9: DNA Oligonucleotides used for Northern Blot detection.

Name	Sequence	Target region
JVO-2405	CCTATGGGAGCGCGGTG	STnc250
JVO-2406	GTCAGAATACGACATTTTGGTACTC	STnc290
JVO-2407	TTATTTGGACTACCTGGATG	STnc340
JVO-2408	TATGAGGAGGACAATTACCG	STnc440
JVO-2445	TACCGGACAATAATCCCTAC	STnc130
JVO-2446	GATAACCTGAGACCCCCCTG	STnc150
JVO-2448	ATATAAACGCGCCAGTCCAT	STnc180
JVO-2466	TCTGGCGGAACCTGCC	STnc220
JVO-2468	CACACCTGTCGGGCGTT	STnc310
JVO-2469	CGCAGTCCCAGGTCAGC	STnc330
JVO-2498	CTTATGTGGGCGTTTTGTTT	STnc350
JVO-2499	AATGACACCAACCTTTTACG	STnc390
JVO-2500	CTAGAGGAGGCGCTAGAAAG	STnc400
JVO-3140	CGGGTGGGATGAAATCGTAA	STnc190
JVO-3141	TTAGTGTCTGGCGAAACGCT	STnc400
JVO-3142	GTTGCTGCGGTGTAATAAGACA	STnc180
JVO-3143	TACGTTTGAGCTCAGGGTCG	STnc180
JVO-3144	TCATGTTACCGGTAAAATACCACC	STnc200
JVO-3249	AGAGAGTCAGCGCCGGG	STnc600
JVO-3250	AATTAAAACCACCCGCCG	STnc620
JVO-3251	CAGGCTACCAACCACCTCC	STnc590
JVO-3252	TATGGAGCGCAACGCC	STnc580
JVO-3253	GCGGTCTGGTGTACCTTCC	STnc610
JVO-3254	CGGGTCATCTTTCAGGCTG	STnc540
JVO-3255	TGCTTATACGCTACCGGGC	STnc560
JVO-3256	CTGCCTAACATCTCGTTTCTCC	STnc570
JVO-3257	GCCACGGTTCTCACCG	STnc480
JVO-3258	CAGCACACTACACAGGGTCG	STnc630
JVO-3259	ACCTTGCTGGCGCTCTC	STnc470
JVO-3260	CATCTTGCGGTCTGGCA	STnc490
JVO-3261	CATCGCGTTGCCAACTT	STnc500
JVO-3262	AAGACCCTGGCGCGGTT	STnc520
JVO-3263	CTTAGCAGCCTTGTAGAAGAGC	STnc640
JVO-3264	AAACTTGACACCGTTCGGC	STnc510
JVO-3265	GTGCCTCCGAACGGAAG	STnc530
JVO-3266	GCGACAATCACGCCCAG	STnc550

10.4. Appendix to Chapter 6

This section contains supplementary Figures and Tables to Chapter 6.

Figure 10.7: Alignment sRNA candidate IG433. The intergenic region between HP1067 and HP1066 as well as 100 bp of the flanking genes from *H. pylori* strain 26695 was aligned with the homologous region of *H. pylori* strain J99 using `Multalign` (Corpet, 1988). Predicted promoter and terminator are indicated in yellow and blue, respectively. Brackets indicate paired bases in the predicted terminator stem-loop. Flanking genes are indicated in grey and their start or stop codons are set in bold. The position of the transcriptional start site according to 5' RACE analysis is indicated by bold nucleotides.

Figure 10.8: Alignment sRNA candidate IG449. The intergenic region between HP1108 and HP1107 as well as 100 bp of the flanking genes from *H. pylori* strain 26695 was aligned with the homologous region of *H. pylori* strain J99 using `Multalign` (Corpet, 1988). Predicted promoter and terminator are indicated in yellow and blue, respectively. Brackets indicate paired bases in the predicted terminator stem-loop. Flanking genes are indicated in grey and their start or stop codons are set in bold.

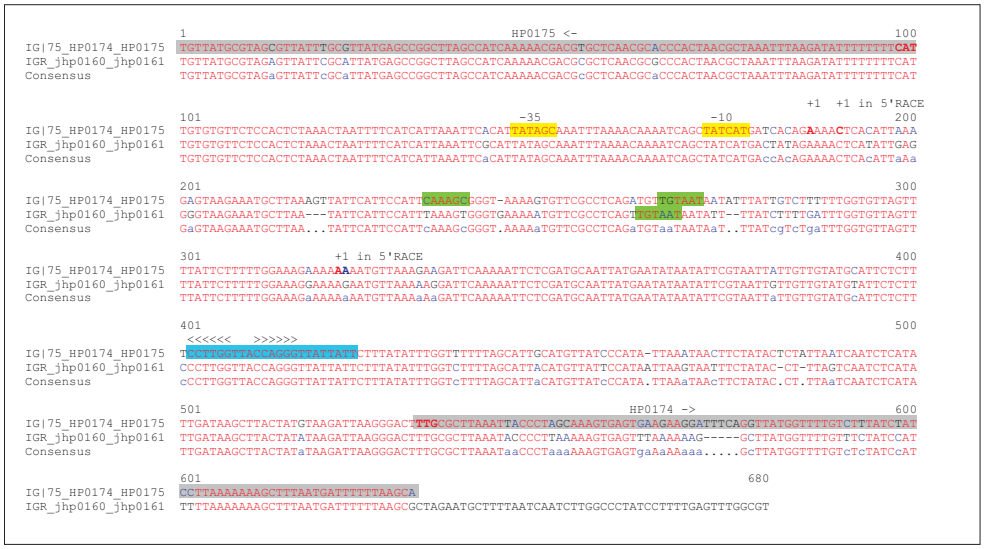


Figure 10.9: Alignment sRNA candidate IG75. The intergenic region between HP0175 and HP0174 as well as 100 bp of the flanking genes from *H. pylori* strain 26695 was aligned with the homologous region of *H. pylori* strain J99 using `Multalign` (Corpet, 1988). The predicted terminator is indicated in blue. Brackets indicate paired bases in the predicted terminator stem-loop. The position of the transcriptional start site according to 5' RACE analysis is indicated by bold nucleotides and a predicted promoter which fits to this start site is indicated in yellow. Alternative predicted promoters are indicated in green. Flanking genes are indicated in grey and their start or stop codons are set in bold.

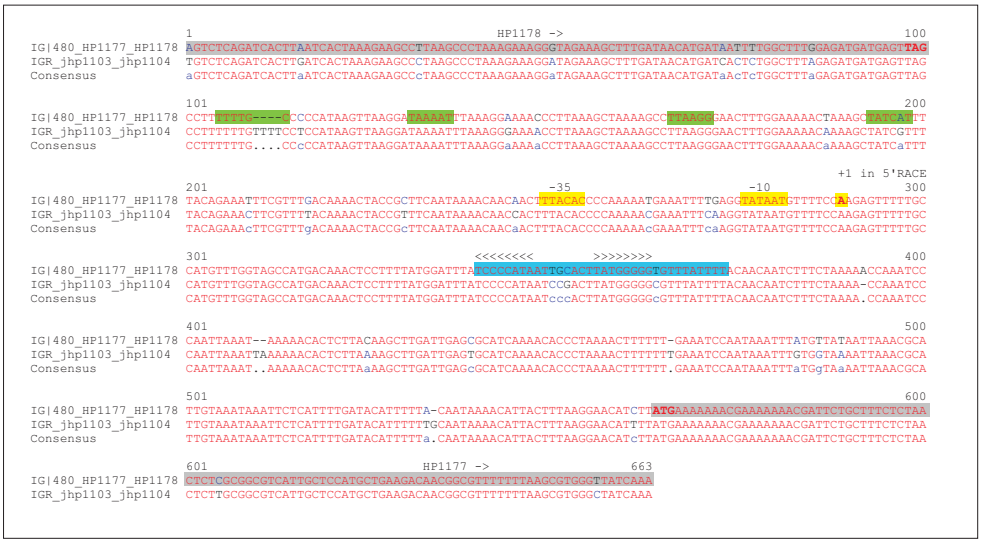


Figure 10.10: Alignment sRNA candidate IG480. The intergenic region between HP1178 and HP1177 as well as 100 bp of the flanking genes from *H. pylori* strain 26695 was aligned with the homologous region of *H. pylori* strain J99 using `Multalign` (Corpet, 1988). The predicted terminator is indicated in blue. Brackets indicate paired bases in the predicted terminator stem-loop. The position of the transcriptional start site according to 5' RACE analysis is indicated by bold nucleotides and a predicted promoter which fits to this start site is indicated in yellow. Alternative predicted promoters are indicated in green. Flanking genes are indicated in grey and their start or stop codons are set in bold.

Figure 10.11: Alignment sRNA candidate IG494. The intergenic region between HP1198 and HP1199 as well as 100 bp of the flanking genes from *H. pylori* strain 26695 was aligned with the homologous region of *H. pylori* strain J99 using `Multalign` (Corpet, 1988). Predicted promoter and terminator are indicated in yellow and blue, respectively. Brackets indicate paired bases in the predicted terminator stem-loop. Flanking genes are indicated in grey and their start or stop codons are set in bold. The position of the transcriptional start site according to 5' RACE analysis is indicated by bold nucleotides.

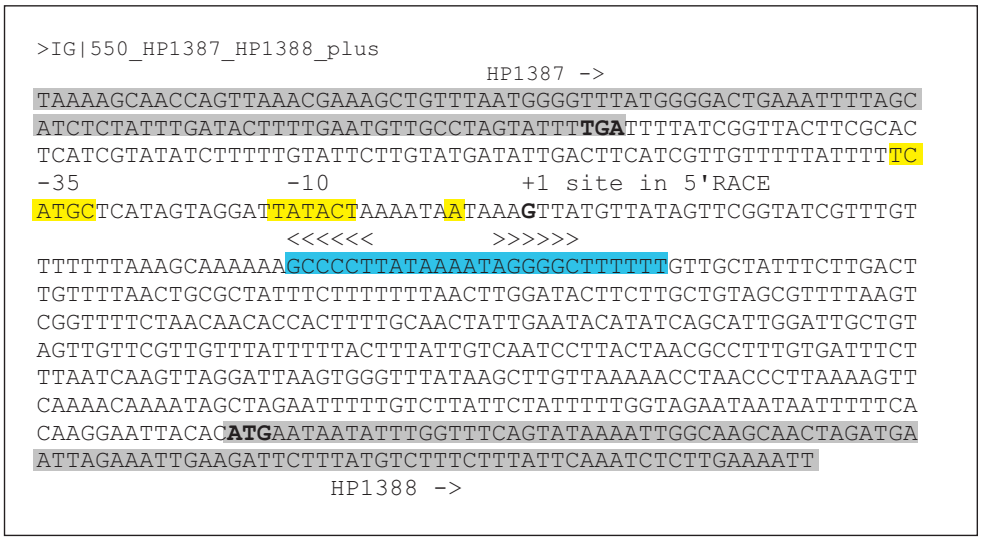


Figure 10.12: Small RNA candidate IG550. Small RNA candidate IG550 is located in the intergenic region between HP1387 and HP1388 of *H. pylori* strain 26695. Predicted promoter and terminator are indicated in yellow and blue, respectively. Brackets indicate paired bases in the predicted terminator stem-loop. The 100 bp of each flanking genes are indicated in grey and their start or stop codons are set in bold, respectively. The position of the transcriptional start site according to 5' RACE analysis is indicated by a bold nucleotide.

Table 10.10: Biocomputationally predicted sRNA candidates in *Helicobacter pylori*.

IGR	Strand	Length	Terminator Score	LFG	RFG	Orientation[c]	Start	End	Distance to LFG	Distance to RFG
IG75 P1	-	240	-3.18	HP0174 (hyp. protein)	HP0175 (cell binding factor 2)	←←→	181542	181781	108	84
IG75 P2[d]	-	146	-3.18	HP0174 (hyp. protein)	HP0175 (cell binding factor 2)	←←→	181542	181687	108	178
IG433	-	101	-4.3	HP1066 (hyp. protein)	HP1067 (*cheY*)	←←→	1126097	1126187	120	81
IG449	-	70	-3.05	HP1107 (*omp23*)	HP1108 (*porC*, *porG*)	←←→	1169977	1170046	99	92
IG480	-	81	-6.31	HP1177 (*omp27*)	HP1178 (*deoD*)	←←←	1245700	1245780	192	184
IG494	+	174	-5.48	HP1198 (*rpoBC*)	HP1199 (*rplL*)	←→←	1277094	1277267	45	6
IG550[e]	+	75	-8.19	HP1387 (DNA polymerase III subunit ϵ)	HP1388 (hyp. protein)	→→→	1449783	1449857	109	331

[a] LFG: left flanking gene.

[b] RFG: right flanking gene.

[c] Orientation of sRNA candidate (middle) and flanking genes (→ and ← denote location of a gene on the clockwise or the counterclockwise strand of the *Helicobacter* chromosome).

[d] For IG75 two promoters were predicted

[e] sRNA candidate IG550 is specific to HP26695

Table 10.11: Experimentally mapped *Helicobacter pylori* promoters based on primer extension or 5'RACE.

Number	Gene	Location	Strand	Distance AUG [bp]	+1 site in *H. pylori* 26695	TSS[a]	Class[b]	Distance [bp][c]	Comment	Reference
HP0011	*hspA (groES)*	9268..9624	-	-59	9683	9683	P	0	σ^{80}	Spohn & Scarlato, 1999b; Suerbaum *et al.*, 1994
HP0071	*ureI*	74747..75334	-	-65	75399	75398	P	1	σ^{80}	Akada *et al.*, 2000; Pflock *et al.*, 2005
HP0073	*ureA*[d]	77240..77956	-	-57	78012	78012	P	0	σ^{70}	Shirai *et al.*, 1999; Spohn & Scarlato, 1999a
HP0088	*rpoD*	92952..94967	-	-72	95039	95040	P	-1	σ^{80}	Beier *et al.*, 1998
HP0097	hyp. prot.	101643..102356	-	-40	102396	102396	P	0	σ^{70}	McGowan *et al.*, 2003
HP0103	*tlbB*	109025..110722	-	-140	110862	110862	P	0	σ^{80}	Delany *et al.*, 2002b
HP0111	*hrcA*	118823..119653	-	-15	119668	119668	P	0	σ^{80}	Spohn & Scarlato, 1999b
HP0115	*flaB*	122948..124492	-	-25	124517	124517	P	0	σ^{54}	Spohn & Scarlato, 1999a
HP0119	hyp. prot.	129383..130768	-	-43	130811	130811	P	0	σ^{80}	Dietz *et al.*, 2002
HP0166	response regulator	173778..174455	-	-67	174522	174520	P	2	σ^{80}	Dietz *et al.*, 2002; Forsyth *et al.*, 2002
HP0220	*nifS*	228339..229502	+	-252	228087	228285	P	-198	?	Pflock *et al.*, 2007
HP0389	*sodB*	398432..399073	-	-21	399094	399094	P	0	σ^{80}	Pesci & Pickett, 1994
HP0407	*bisC*	419077..421467	+	-34	419043	419043	P	0	σ^{70}	McGowan *et al.*, 2003
HP0427	hyp. prot.	444265..444600	-	2	444598	444600	P	-2	σ^{80}	Dietz *et al.*, 2002; Müller *et al.*, 2006; Pflock *et al.*, 2004
-	*cagB*[e] P2	578740..579087	-	P2:-581	579668	579669	O	-1	σ^{70}	Spohn *et al.*, 1997,
HP0547	*cagA*	579921..583481	+	P1:-104	579817	579817	P	0	σ^{70}	Spohn *et al.*, 1997
HP0578	hyp. prot.	608300..610336	-	-58	610394	-	P	-	?	McGowan *et al.*, 2003

continued on next page

Number	Gene	Location	Strand	Distance AUG [bp]	+1 site in *H. pylori* 26695	TSS[a]	Class[b]	Distance [bp][c]	Comment	Reference
HP0600	*spaB*	635337..637118	-	-27	637145	637146	P	-1	σ^{70}	McGowan *et al.*, 2003
HP0601	*flaA*[d]	637282..638814	+	-50	637232	637232	P	0	σ^{28} (P-50)	Leying *et al.*, 1992; McGowan *et al.*, 2003; Spohn & Scarlato, 1999a
HP0653	*pfr*	698770..699273	-	-29	699302	699301	P	1	σ^{80}	Delany *et al.*, 2001b
HP0654	hyp. prot.	699570..700652	+	-26	699544	699544	P	0	σ^{80}	Delany *et al.*, 2001b
HP0682	hyp. prot.	732133..732513	-	-6	732519	732519	P	0	σ^{80}	Forsyth *et al.*, 2002
HP0690	*fadA*	740559..741734	+	-106	740453	740534	P	-81	?	Pflock *et al.*, 2007
HP0695	*hyuA*	745801..747942	+	-39	745762	745763	P	-1	σ^{80}	Pflock *et al.*, 2007
HP0698	hyp. prot. - *flgR*[f]	750956..751045 (HP0698)	+	-22	750934	750935	P	-1	σ^{80}	Spohn & Scarlato, 1999a
HP0797	*hpaA*	965177..966724	-	-80 (-72)	854819	854817	P	2	σ^{70}(P-80),	Jones *et al.*, 1997; McGowan *et al.*, 2003
HP0870	*flgE*	920417..922573	-	-30	922603	922604	P	-1	σ^{54}	Spohn & Scarlato, 1999a
HP0875	*katA*	925571..927088	-	-55	927143	927143	P	0	σ^{80}	Delany *et al.*, 2001a
HP0876	*frpB*	927411..929786	+	-78	927333	927332	P	1	σ^{80}	Delany *et al.*, 2001a
HP0878	hyp. prot.	930391..930564	+	-10	930313	930313	P	0	σ^{70}	McGowan *et al.*, 2003
HP0906	hyp. prot.	956485..958068	+	-45	956440	-	P	-	σ^{54}	Spohn & Scarlato, 1999a
HP0912	*hopC*	637282..638814	+	-105	965072	965072	P	0	σ^{70}	McGowan *et al.*, 2003; Odenbreit *et al.*, 1999
HP1010	*ppk*	1072429..1074456	+	-43	1072386	-	P	-	σ^{70}	McGowan *et al.*, 2003

continued on next page

Number	Gene	Location	Strand	Distance AUG [bp]	+1 site in *H. pylori* 26695	TSS[a]	Class[b]	Distance [bp][c]	Comment	Reference
HP1018-HP1019	hyp. prot.-*htrA*	1081440..1081586	+	-42	1081398	1081400	P	-2	σ^{80}	Pflock *et al.*, 2004
HP1024	*cbpA*	1087633..1088499	+	-27	1087606	1087516	P	90	σ^{80}	Spohn & Scarlato, 1999b
HP1027	*fur*	1090212..1090664	+	-41	1090171	1090171	P	0	σ^{80}	Delany *et al.*, 2002a
HP1041	*flbA (flhA)*	1100927..1103128	+	-64	1100863	1100883	P	-20	σ^{80}	Schmitz *et al.*, 1997
HP1043	response regulator	1104745..1105416	-	-34	1105450	1105449	P	1	σ^{80}	Delany *et al.*, 2002b
HP1067	*cheY*	1126268..1126642	+	-84	1126184	1126183	P	1	σ^{80}	Beier *et al.*, 1997
HP1120	hyp. prot....*flgK*	1186442..1186876	-	-24	1186900	-	P	-	σ^{54}	Spohn & Scarlato, 1999a
HP1139	*soj*	1200639..1201433	-	-117	1201550	1201551	S	-1	σ^{70}	McGowan *et al.*, 2003
HP1186	carbonic anhydrase	1255772..1256380	+	-43	1255729	1255729	P	0	σ^{80}	Wen *et al.*, 2007
HP1260	*nuoA*	1333813..1334214	+	-38	1333775	1333777	P	-2	σ^{70}	McGowan *et al.*, 2003
HP1335	*trmU*	1395894..1396976	-	-24	1397000	1396947	P	53	σ^{70}	McGowan *et al.*, 2003
HP1362	*dnaB*	1422915..1424381	-	-154	1424535	1424702	P	-167	σ^{54}	McGowan *et al.*, 2003
HP1400	*fecA3*	1461431..1463959	+	-113	1461318	1461028	P	290	σ^{80}	Ernst *et al.*, 2006
HP1408	hyp. prot.	1477542..1477877	+	2	1477544	1477542	P	2	σ^{80}	Dietz *et al.*, 2002; Müller *et al.*, 2006; Pflock *et al.*, 2004
HP1423	*orf03*	1494708..1494962	-	0	1494962	1494962	P	0	σ^{80}	Porwollik *et al.*, 1999
HP1432	histidine and glutamine-rich protein	1502586..1502804	+	-44	1502542	1502550	P	-8	σ^{80}	Forsyth *et al.*, 2002

continued on next page

Number	Gene	Location	Strand	Distance AUG [bp]	+1 site in *H. pylori* 26695	TSS[a]	Class[b]	Distance [bp][c]	Comment	Reference
HP1494	*murE*	1567157..1568500	+	-39	1567118	-	P	-	σ^{70}	McGowan *et al.*, 2003
HP1512	*frpB4*	1584447..1587080	+	-55	1584392	1584392	P	0	σ^{80}	Ernst *et al.*, 2006
HP1559	*flgBC*	1640954..1641376	-	-25	1641401	1641505	P	-104	σ^{54}	Spohn & Scarlato, 1999a
HP1562	*ceuE*	1643982..1644983	-	-25	1645008	1645007	P	1	σ^{80}	Delany *et al.*, 2001a
HP1563	*tsaA*	1645224..1645820	+	-96	1645128	1645128	P	0	σ^{80}	Delany *et al.*, 2001a
HPnc6350	asRNA A	1245700.. 1245780	-	-	1245780	1245780		0	σ^{80}	5' RACE, this study
HPnc6320	asRNA B	1243405.. 1243474	-	-	1243474	1243474		0	σ^{80}	5' RACE, this study
HPnc7630	asRNA C	1503081.. 1503160	-	-	1503160	1503160		0	σ^{80}	5' RACE, this study
HPnc8170	asRNA D	1612518.. 1612596	-	-	1612596	1612596		0	σ^{80}	5' RACE, this study
HPnc0040	asRNA E	22856.. 22931	-	-	22931	22931		0	σ^{80}	5' RACE, this study
HPnc8060	asRNA F	1589890.. 1589984	-	-	1589984	1589984		0	σ^{80}	5' RACE, this study
HPnc7620	spRNA C	1503081..1503361	+	-	1503081	1503081		0	σ^{80}	5' RACE, this study
HPnc8160	spRNA D	1612518..1612827	+	-	1612518	1612518		0	σ^{80}	5' RACE, this study
HPnc6670	sRNA B	1307822..1307963	-	-	1307963	1307963		0	σ^{80}	5' RACE, this study
HPnc2090	sRNA C1	479770..479856	-	-	479856	479856		0	σ^{80}	5' RACE, this study
HPnc2640	sRNA D	568309..568522	-	-	568522	568522		0	σ^{80}	5' RACE, this study
HPnc2420	sRNA H	537522..537624	-	-	537624	537624		0	σ^{80}	5' RACE, this study

continued on next page

Number	Gene	Location	Strand	Distance AUG [bp]	+1 site in *H. pylori* 26695	TSS[a]	Class[b]	Distance [bp][c]	Comment	Reference
HPnc4590	sRNA I	964751..964802	+	-	964751	964751		0	σ^{80}	5' RACE, this study
HPnc7830	sRNA MP1	1524441.. 1524681	-	-	1524681	1524681		0	σ^{80}	5' RACE, this study
HPnc6160	asRNA 3	1217306..1217378	+	-	1217306	1217306		0	σ^{80}	5' RACE, this study
HPnc4870	asRNA 5	998717..998848	+	-	998717	998717		0		σ^{80}5' RACE, this study
HPnc7430	asRNA 10	1470865..1470983	-	-	1470983	1470983		0		σ^{80}5' RACE, this study
HPnc1880	asRNA 11	445011.. 445139	-	-	445139	445139		0	σ^{80}	5' RACE, this study
HPnc1810	ssRNA IIIa	438178.. 438908	-	-	438908	438908		0	σ^{80}	5' RACE, this study
HPr01	23S rRNA	445248..448223	+	-	444979	444979	P	0	σ^{80}	5' RACE, this study
HPr06	23S rRNA	1473917..1476893	-	-	1477163	1477163	P	0	σ^{80}	5' RACE, this study

[a] Transcriptional start site (TSS) mapped in *H. pylori* 26695 based on 454 deep sequencing data.

[b] TSS class according to the manual annotation described in Section 6.1.5 in Chapter 6. P: primary, S: secondary, and O: orphan.

[c] Distance between the transcriptional start site described in the literature or mapped by 5' RACE and the TSS based on the 454 data: $TSS_{literature/5'RACE} - TSS_{454}$.

[d] Promoters for *ureA* and *flaA* were confirmed by 5' RACE

[e] The *cagB* gene was identified in strain G27 but is not annotated in *H. pylori* 26695.

[f] The *flgR* promoter is located upstream of HP0698.

Table 10.12: Read distribution on annotations.

Library	Total reads	<12 nt	No match in HP 26695	23S rRNA	16S rRNA	5S rRNA	tRNA	tmRNA	RNase P RNA	SRP RNA	mRNA	as mRNA	IGR	as rRNAs, hkRNAs and tRNAs
C-	528373	7515	75888	224307	101306	5411	38364	1998	1169	1282	60782	3272	7046	33
C+	528169	14984	90257	81177	33560	18794	142788	1598	8849	5577	115119	6164	9246	56
AS-	427455	8265	63906	52012	42930	8634	45326	3683	2482	2683	169410	13383	14554	187
AS+	540133	18019	108421	30899	9849	13935	133956	1525	8174	4868	159737	24203	26263	284
PL-	268841	5130	61008	49677	37018	4396	47628	1242	893	1168	51943	3790	4880	68
PL+	315309	6602	57064	39014	14553	20032	89390	844	3248	2635	68218	5768	7838	103
AGS-	280713	4547	93030	44836	35744	4681	40998	982	696	876	44424	4498	5335	66
AGS+	223705	5280	78117	18332	6510	10414	57476	375	1633	1842	35184	3881	4582	79
Huh7-	266621	865	66905	58946	38699	4858	8090	1076	533	205	75632	5415	5362	35
HuH7-	308759	657	100705	52265	12699	10927	20300	1313	2396	936	88559	11114	6751	137

Table 10.13: Leaderless mRNAs in *Helicobacter pylori*. Genes that turned out to be leaderless during transcriptional start site annotation in *H. pylori* 26695 based on deep sequencing data. Location indicates start and end of the ORF in the *H. pylori* 2695 genome. "_R" following a gene name indicates that this gene was reannotated based on the 454 data and conservation of the start codon in other *Helicobacter* strains.

Gene	Location of AUG start codon	Description
HP1529	1608997 - 1608995	*dnaA*; chromosomal replication initiation protein
HP1139	1201433 - 1201431	SpoOJ regulator (*soj*) chromosome partitioning protein
HP0925	988604 - 988606	recombination protein RecR
HP0376	384259 - 384261	ferrochelatase *hemH*
HP0929	991805 - 991807	geranyltranstransferase (IspA)
HP0413	426838 - 426836	putative transposase
HP0414	426876 - 426878	IS200 insertion sequence from SARA17
HP1008	1069967 - 1069969	IS200 insertion sequence from SARA17
HP1181	1249488 - 1249490	multidrug-efflux transporter
HP1183	1252759 - 1252757	Na+/H+ antiporter (*napA*)
HP1216	1293586 - 1293584	organic solvent tolerance protein
HP0818_R	870451 - 870453	osmoprotection protein (*proWX*)
HP0498	524081 - 524083	sodium- and chloride-dependent transporter; neurotransmitter:Na+ symporter
HP1365	1427604 - 1427602	response regulator, OmpR family
HP0329	345305 - 345303	NH(3)-dependent NAD+ synthetase (*nadE*)
HP1394	1455819 - 1455817	hypothetical protein NAD+ kinase
HP0112_R	120046 - 120048	hypothetical protein
HP0427	444600 - 444598	hypothetical protein
HP0820	872931 - 872933	hypothetical protein
HP1408	1477542 - 1477544	hypothetical protein
HP1423	1494962 - 1494960	hypothetical protein
HP0151_R	161202 - 161204	hypothetical protein
HP0806	860357 - 860359	hypothetical protein
HP0897_R	950133 - 950135	hypothetical protein
HP1007	1069927 - 1069929	frameshift gene

Table 10.14: DNA Oligonucleotides used for Northern Blot detection. Sequences are given in 5' → 3' direction.

Gene	Name	Oligo	Sequence
rnpB	RNase P RNA	JVO-0210	CGAAGCGTGTATCAATTTAGAC
ffs	SRP RNA	JVO-0211	GGGACTCTGCTGTATTCCTAC
tmRNA	tmRNA	JVO-0212	CTGGAGCGTAATCTGTGTTG
5S rRNA	5S rRNA	JVO-0485	TCGGAATGGTTAACTGGGTAGTTCCT
23S rRNA	23S rRNA	JVO-0586	GCATAGCTTATCGCAGTCTAGT
HPnc6350	asRNA A	JVO-0231	GAGTTTGTCATGGCTACCAA
HPnc6320	asRNA B	JVO-0513	GCCATGGAAAATTAAAAATG
HPnc7630	asRNA C	JVO-0514	CATGCCATGAAACACAAAAG
HPnc8170	asRNA D	JVO-0548	GCAGACCAACATTGCA
HPnc8060	asRNA F	JVO-0550	CTAATTTTTATTCCACTAGAGATTA
HPnc7620	spRNA C	JVO-2303	CCTTTTGACATAGGATTTGC
HPnc8160	spRNA D	JVO-2135	GATCGCATGGCATGCT
HPnc6561	6S RNA	JVO-2136	AACACGAATCATCTAGGCGAT
HPnc5490	sRNA A	JVO-2134	AAACCATAAGGAATGGTTGGAT
HPnc6670	sRNA B	JVO-2621	AATGCTGAAGCTTCTAGAATGAT
HPnc2630	sRNA D	JVO-2623	GATTTGTTTGTTTATGCCAAA
HPnc2240/ HPnc6000	asRNA 1a/b	JVO-2133	GTGAACCATAGGTTGAGTTCCTATAG
HPnc2450	asRNA 7	JVO-2635	CGAGAAATACCTCCACACAAT
HPnc1880/ HPnc7450	asRNA 11	JVO-2702	GCGTTATAAAAAGATTAGGGATCA
HPnc1470	sRNA L	JVO-2708	TTAAACTTTAACAACTCTTTAATTTTCAA
HPnc3320	ssRNA I	JVO-2627	CTCATTTGTACATCCGCTTTA
HPnc1810/ HPnc7520	ssRNA IIIa/b	JVO-2707	CTATCCTCTTTCTCTTTAGGAGTTG
HPnc4160	asRNA G	JVO-2624	GTTGATTAAAATGCTAAGTTATAGTAAAGA
HPnc4170	spRNA G	JVO-2625	CTGACGCTCTTACCTTAATTGA
HPnc2090/ HPnc5320	sRNA C1/2	JVO-2622	AGAAAAGGAGATAACCTAACATGA
HPnc4590	sRNA I	JVO-2704	AAAGGAGATAACCAACTATGAAGTT
invR	*Salmonella* InvR	JVO-0222	GATAAATGCAACGTAAGAGACAAATG
sraH	*Salmonella* SraH	JVO-0157	GGGTGCGCGAATACTG
rrf	*Salmonella* 5S rRNA	JVO-0322	CTACGGCGTTTCACTTCTGAGTTC

Table 10.15: DNA Oligonucleotides used for 5' RACE analysis. Sequences are given in 5' → 3'direction.

Gene	Name	Oligo	Sequence
IG75	IG75	JVO-0226	AGAATGCATACAACAATAATTACG
IG433	IG433	JVO-0240	GAACGCTCCATTTA
IG550	IG550	JVO-0245	GGGGCTTTTTTG
HP0601	*flaA*	JVO-0974	TCAATCGCTCCAATGAAGTT
HP0073	*ureA*	JVO-0214	TACCGCTTCTACATAGTTAAGCTT
HPnc6350	asRNA A (IG480)	JVO-0231	GAGTTTGTCATGGCTACCAA
HPnc6320	asRNA B	JVO-0794	ATAAATTCTAAAAAGGAGTTTGCCA
HPnc7630	asRNA C	JVO-0514	CATGCCATGAAACACAAAAG
HPnc8170	asRNA D	JVO-0548	GCAGACCAACATTGCA
HPnc0040	asRNA E	JVO-0549	GCCTCATAGTTAGGATATGG
HPnc8060	asRNA F	JVO-0795	TCTAGGAGACTTCTATGAGAAAAAATC
HPnc7620	spRNA C	JVO-2304	CTTTTGTGTTTCATGGCATG
HPnc8160	spRNA D	JVO-2135	GATCGCATGGCATGCT
HPnc6670	sRNA B	JVO-2621	AATGCTGAAGCTTCTAGAATGAT
HPnc2090	sRNA C1	JVO-2622	AGAAAAGGAGATAACCTAACATGA
HPnc2640	sRNA D	JVO-2623	GATTTGTTTGTTTATGCCAAA
HPnc2420	sRNA H	JVO-2703	ACACAAGGCAAGTGTGATAAAC
HPnc4590	sRNA I	JVO-2704	AAAGGAGATAACCAACTATGAAGTT
HPnc7830	sRNA MP1	JVO-2709	CTCTCACGCATCATATCTATAAAG
HPnc6160	asRNA 3	JVO-2631	CTAATGTGACCGGTGTGTTG
HPnc4870	asRNA 5	JVO-2633	AAGAACAAGCCCTAAAATTTGT
HPnc7430	asRNA 10	JVO-2698	TGGGAATAAAGACTTGGAAAATTTAAGT
HPnc1880	asRNA 11	JVO-2702	GCGTTATAAAAAGATTAGGGATCA
HPnc1810	ssRNA IIIa	JVO-2707	CTATCCTCTTTCTCTTTAGGAGTTG
HPr01	23S rRNA	JVO-2741	AGCTTTTAGCTTGTAGAACTTGCTT
HPr06	23S rRNA	JVO-2741	AGCTTTTAGCTTGTAGAACTTGCTT