Small RNAs, Big Consequences

Post-transcriptional Regulation and Adaptive Immunity in Bacteria

MIRTHE HOEKZEMA
Abstract

It is nowadays widely accepted that non-coding RNAs play important roles in post-transcriptional regulation of genes in all kingdoms of life. In bacteria, the largest group of RNA regulators are the small RNAs (sRNAs). Almost all sRNAs act through anti-sense base-pairing with target mRNAs, and by doing so regulate their translation and/or stability. As important modulators of gene expression, sRNAs are involved in all aspects of bacterial physiology. My studies aimed to deepen our understanding of the mechanisms behind sRNA-mediated gene regulation. We have shown that translation of the di-guanylate-cyclase YdaM, a major player in the biofilm regulatory cascade, is repressed by the sRNAs OmrA and OmrB. OmrAB require the RNA chaperone protein Hfq for efficient regulation. Interestingly, our results suggest a non-canonical mechanism for Hfq-mediated ydaM-OmrA/B base-pairing. Instead of serving as RNA interaction platform, Hfq restructures the ydaM mRNA to enable sRNA binding.

We also addressed the question of how bacteria utilize regulatory RNAs to create phenotypic heterogeneity by studying the role of the tisB/istR-1 type 1 toxin-antitoxin system in SOS-induced persister cell formation in E. coli.

In addition, I have investigated the prokaryotic CRISPR-Cas immune system, which has led to the development of two molecular tools. The CRISPR-Cas adaptive immune system consists of a CRISPR array, where palindromic repeats are interspaced by unique spacer sequences derived from foreign genetic elements, and the CRISPR-associated (Cas) proteins. In the adaptation stage, memory is created by insertion of spacer sequences into the CRISPR array. We developed a fluorescent reporter that accurately and sensitively detects spacer integration events (denoted: “acquisition”) in single cells and in real-time. In the effector stage of immunity, crRNAs, consisting of one spacer-repeat unit, associate with the Cas proteins to form a ribonucleoprotein complex that surveys the cell for invader DNA. Target identification occurs by base-pairing between the crRNA and the complementary sequence in the target nucleic acid, which triggers degradation. We repurposed the E. coli type I-E CRISPR-Cas effector complex Cascade for specific reprogrammable transcriptional gene silencing.

The studies presented herein thus contributes to our understanding of RNA-based target identification for gene regulation and adaptive immunity.

Keywords: small RNA, sRNA, non-coding RNA, Hfq, gene regulation, post-transcriptional regulation, biofilm, OmrA, OmrB, toxin-antitoxin, TisB, IstR-1, Persisters, CRISPR-Cas, CRISPR-Cas adaptation reporter, Escherichia coli

Mirthe Hoekzema, Department of Cell and Molecular Biology, Box 596, Uppsala University, SE-75124 Uppsala, Sweden.

© Mirthe Hoekzema 2018

ISSN 1651-6214
urn:nbn:se:uu:diva-343526 (http://urn.kb.se/resolve?urn:nbn:se:uu:diva-343526)
Many of the truths we cling to depend greatly on our own point of view.

Obi-Wan Kenobi -

Voor mijn vader
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


*These authors contributed equally to this work

Reprints were made with permission from the respective publishers.
Contents

Introduction ............................................................................................... 11
Gene regulation .......................................................................................... 11
Regulatory RNAs .................................................................................... 12
sRNA biogenesis .................................................................................. 14
sRNA modes of action ........................................................................ 16
The RNA chaperone Hfq ....................................................................... 20
Advantages of regulating with sRNAs .................................................. 23
The OmrA and OmrB sRNAs ............................................................. 25
sRNAs in motility and biofilm formation ........................................... 26
Toxin-antitoxin systems ........................................................................ 28
Type I toxin-antitoxin systems ............................................................ 28
The tisB/istR-I toxin-antitoxin system ................................................ 29
Role in persister formation .................................................................. 29
The CRISPR-Cas adaptive immune system ......................................... 32
Diversity and classification ................................................................. 33
Adaptation .............................................................................................. 34
Expression .............................................................................................. 38
Interference ............................................................................................ 39
CRISPR-Cas applications ................................................................... 41

Present investigation ................................................................................ 44
Paper I .................................................................................................. 44
Paper II ................................................................................................ 48
Paper III ................................................................................................ 51
Paper IV ................................................................................................ 55

Concluding remarks ................................................................................ 58
Sammanfattning på svenska .................................................................. 59
Samenvatting in het Nederlands ............................................................... 62
Acknowledgements ............................................................................... 65
References ............................................................................................... 68
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bfp</td>
<td>Blue fluorescent protein</td>
</tr>
<tr>
<td>bp</td>
<td>base-pair</td>
</tr>
<tr>
<td>cas</td>
<td>CRISPR-associated protein</td>
</tr>
<tr>
<td>Cascade</td>
<td>CRISPR-associated complex for antiviral defense</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>crRNA</td>
<td>CRISPR-RNA</td>
</tr>
<tr>
<td>CTD</td>
<td>C-terminal domain</td>
</tr>
<tr>
<td>DGC</td>
<td>Diguanylate cyclase</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-stranded DNA</td>
</tr>
<tr>
<td>Gfp</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>IGR</td>
<td>Intergenic region</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PAM</td>
<td>Protospacer adjacent motif</td>
</tr>
<tr>
<td>PDE</td>
<td>Phosphodiesterase</td>
</tr>
<tr>
<td>pre-crRNA</td>
<td>Precursor CRISPR RNA</td>
</tr>
<tr>
<td>RBP</td>
<td>RNA binding protein</td>
</tr>
<tr>
<td>RBS</td>
<td>Ribosome binding site</td>
</tr>
<tr>
<td>SD</td>
<td>Shine-Dalgarno sequence</td>
</tr>
<tr>
<td>sgRNA</td>
<td>single guide RNA</td>
</tr>
<tr>
<td>sRNA</td>
<td>small RNA</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single-stranded DNA</td>
</tr>
<tr>
<td>TA</td>
<td>Toxin-antitoxin</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>tracrRNA</td>
<td><em>trans</em>-activating CRISPR RNA</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>Yfp</td>
<td>Yellow fluorescent protein</td>
</tr>
</tbody>
</table>
Introduction

All organisms interact with their environment. This interaction is necessary to gain sustenance, but it also poses a threat. Bacteria have therefore developed a tremendous capability to sense their environment and rapidly respond to changes by adjusting their gene expression. This rewiring is governed by complex regulatory networks, and bacterial small RNAs (sRNAs), the topic of papers I and II, are an important class of regulators. Another threat when interacting with the environment is that of foreign agents infiltrating the system. For us humans these are for example viruses and bacteria, and for bacteria themselves, they are bacteriophages and other mobile genetic elements. Much like we have our immune system, bacteria have specialized mechanisms to deal with these invaders. One of these systems, the topic of papers III and IV, is the CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins) adaptive immune system. While post-transcriptional regulation and adaptive immunity might seem like two unrelated topics, a common denominator is the use of base-pairing RNAs, sRNAs and CRISPR RNAs (crRNAs), respectively, to specifically identify target RNA or DNA within the large pool of nucleic acids present in the cell. Thus, one can consider (small) RNA biology as the overarching theme of this thesis. The work included in the thesis uses the Gram-negative model organism Escherichia coli. Therefore, the introduction will mainly focus on what we know about sRNAs as well as CRISPR-Cas in E. coli and closely related enterobacteria. It should be noted that sRNAs as well as CRISPR-Cas also play important roles in other species, and that while the general theme is the same, there are substantial and important differences.

Gene regulation

All cells in our body contain the same genetic information, yet they are capable of forming the many different cell types our anatomy is composed of. Individual cells of a bacterial species also behave differently, depending on the circumstances they find themselves in. This is due to differential gene expression, which is the result of the regulation. Gene expression is the process by which the information from a gene is used to generate a functional product, usually in the form of a protein (Figure 1). The first step is to transcribe, or copy, the coding part of the DNA into a RNA molecule. If the gene contains
information to generate a protein, the RNA is called messenger RNA (mRNA) and is used as a template by a complex molecular machine called ribosome to produce proteins in a process termed translation. Gene regulation can occur at all levels of expression. Regulation at the level of mRNA production is transcriptional regulation, and usually depends on DNA binding proteins called transcription factors (TFs). Transcriptional regulation can be both positive or negative depending on if the TF is an activator or repressor. Post-transcriptional regulation occurs after a mRNA is produced, but before it is translated. These are processes that control the rate at which a mRNA is translated into protein, and/or its rate of degradation. Even after a mRNA is translated, regulation can take place at the level of protein stability or activity. Examples of post-translational regulation involve modifications such as phosphorylation and ubiquitination, or sequestration.

Regulatory RNAs

It is nowadays widely accepted that RNAs play an important role in post-transcriptional regulation in all kingdoms of life. However, for a long time, RNA was considered a largely inert molecule, merely supporting the protein syn-
thesis machinery as information carrier (mRNA), scaffold for ribosome assembly (ribosomal RNA), and adapter molecule translating the genetic code from nucleotides into the amino acids of a protein (transfer RNA). It was not until the 1980s that it was demonstrated that antisense RNAs could be bona fide regulators; they were shown to control plasmid copy number (Tomizawa et al., 1981; Stougaard et al., 1981). Remarkably, the possibility that RNA can regulate gene expression was suggested even as early as 1961 (Jacob and Monod, 1961) but was largely ignored for decades. After the discovery of antisense RNAs on plasmids, it took until 1984 before the first chromosomally encoded regulatory RNA, MicF, was characterized in E. coli (Mizuno et al., 1984). Fast-forward a few more years, until 1994, and we saw the discovery of the first eukaryotic regulatory RNA, the microRNA lin-4 in Caenorhabditis elegans (Lee et al., 1993). Subsequent systematic searches have revealed the presence of regulatory RNAs in almost all phyla of bacteria and many archaeal species (Livny and Waldor, 2007; Babski et al., 2014; Wagner and Romby, 2015). Similarly, microRNAs have been found in many more animals as well as in plants (Bartel, 2004; Voinnet, 2009).

The largest group of bacterial RNA regulators, the small RNAs (sRNAs), act through antisense base-pairing with their target mRNAs. By doing so, they regulate translation and/or mRNA stability (Waters and Storz, 2009; Wagner and Romby, 2015) (see ‘sRNA modes of action’). While the eukaryotic microRNAs act as guides for effector proteins (in the RISC complex), sRNAs usually are the primary regulators, although many require association with the RNA chaperone Hfq for stability and function (see ‘The RNA chaperone Hfq’). sRNAs are heterologous in structure and length, roughly between 50 and 300 nt. A typical sRNA consists of three functional elements; the relatively unstructured and conserved ‘seed-sequence’ is used for base-pairing with target mRNAs (Papenfort et al., 2010), often followed by an A/U rich Hfq binding domain and a U-rich Rho-independent terminator stemloop (Figure 2). The majority of sRNAs are non-coding, though there are exceptions where the sRNA not only act as an antisense regulator but also codes for a small peptide (reviewed in (Gimpel and Brantl, 2017)). sRNAs are important modulators of gene expression, and involved in all aspects of bacterial physiology such as stress responses, virulence, and motility/sessility decisions (reviewed in (Romby et al., 2006; Mika and Hengge, 2013; Papenfort and Vogel, 2014; Holmqvist and Wagner, 2017)).
Figure 2. Schematic depiction of the general features of trans-encoded Hfq-associated sRNAs. The highly-conserved seed region is used for base-pairing with target mRNAs. At the 3’ end the typical stemloop and U-stretch of a Rho-independent terminator, which is used for interaction with Hfq (indicated by grey circles). Internal A/U-rich sequences form an additional Hfq binding domain.

Other examples of RNA regulators in bacteria are riboswitches, where the 5’ untranslated region (UTR) of an mRNA is folded in a complex structure that can interact with, and change conformation upon, binding of small molecules (reviewed in (Mellin and Cossart, 2015; Sherwood and Henkin, 2016)). These alternate structures usually form or disrupt transcriptional (anti-) terminators, or cover/ uncover ribosome binding sites (RBS). Similarly, RNA thermometers are structures that react to changes in temperature (reviewed in (Kortmann and Narberhaus, 2012)). Both of these are examples of cis-acting RNA regulatory elements, in contrast to the previously discussed sRNAs with act in trans. The remainder of this thesis will be focused around various aspects of sRNAs and crRNAs.

sRNA biogenesis

sRNAs come in two flavors, cis-encoded and trans-encoded (Figure 3). Cis-encoded sRNAs arise from transcription of the opposite strand of an mRNA-encoding locus, and consequently have extensive complementarity to the target mRNA. They can therefore readily form stable duplexes, however full duplex formation is often not required for activity (Brantl, 2007; Georg and Hess, 2011). This type of sRNA is mostly found on plasmids where they control copy number, as well as on other mobile genetic elements. Some reside on the chromosome, often encoding antitoxin RNAs of type I toxin-antitoxin systems (Wagner et al., 2002; Brantl, 2007; Bloch et al., 2017) (see also ‘Toxin-antitoxin systems’). It should be noted transcriptomics data has revealed that antisense transcription is prevalent in many bacteria ((Georg and Hess, 2011; Lasa et al., 2012; Wade and Grainger, 2014) and references therein), yet it remains a matter of debate how many, and to what extent these antisense transcripts are functional (Wade and Grainger, 2014).
Figure 3. Schematic representation of the different sRNA biogenesis pathways. Cis-encoded sRNAs are transcribed from the same locus, but opposite strand, as their mRNA target. Trans-encoded sRNAs are transcribed from independent transcription units in the IGR, but can also originate from coding sequences, either from an mRNA-internal promoter or through processing of the mRNA.

In contrast, trans-encoded RNAs originate from a different locus, and therefore share limited complementarity with, their mRNA targets. Due to this limited complementarity, trans-encoded sRNAs can base-pair with multiple mRNA targets. However, unlike cis-encoded sRNAs, enterobacterial trans-encoded sRNAs often require the aid of the RNA chaperone Hfq to form a stable duplex (see also ‘The RNA chaperone Hfq’). Some sRNAs require processing of the primary transcript to become functional. For example, the 56 nucleotide (nt) long ArcZ sRNA is released from the 118 nt long precursor pre-ArcZ by the endoribonuclease RNase E (Argaman et al., 2001; Papenfort et al., 2009; Chao et al., 2017). The first few serendipitously found trans-encoded sRNAs originated from independent transcription units. I.e., they have their own promoter and terminator, and their genes are located in intergenic regions (IGRs), between protein-coding genes (reviewed in (Wassarman et al., 1999)). Subsequent bioinformatics searches used these characteristics (promoter and terminator sequences spaced reasonably close in the IGRs), and sometimes also taking sequence conservation in consideration, to identify many more such sRNAs (Argaman et al., 2001; Wassarman et al., 2001; Rivas et al., 2001). The drawback of such approaches is that one will find what one was looking for, that is, sRNAs originating in IGRs. With the arrival of more unbiased large scale RNA-sequencing approaches came an appreciation that
sRNAs could originate from other genomic locations, such as the 5’ UTRs (Loh et al., 2009; DebRoy et al., 2014; Mellin et al., 2014) and 3’ UTRs (reviewed in (Miyakoshi, Chao, and Vogel, 2015a)) of mRNAs. 3’UTR-encoded sRNAs come in two types, those transcribed from a promoter located inside the open reading frame (ORF) of a gene, and those derived from processing of the mRNA (Miyakoshi, Chao, and Vogel, 2015a) (Figure 3).

sRNA modes of action
Most sRNAs act through sequence-specific base-pairing with their mRNA targets to affect translation and/or stability, either negatively or positively (Waters and Storz, 2009; Wagner and Romby, 2015). An sRNA often has several targets and, while it represses one target, it can activate another. The outcome depends on the characteristics of the interaction, such as the location of the binding site on the target mRNA, and whether or not accessory proteins are employed. There is a wide range of mechanisms by which sRNA affect gene expression. Several are common, others less common, as detailed below (See Figures 4, and 5).

Inhibition by competition with initiating ribosomes
Translation begins with assembly of the initiation complex which, in bacteria, consists of the 30S ribosomal subunit, tRNA<sub>FMet</sub> and initiation factors, on the ribosome binding site (RBS) of the mRNA (reviewed in (Simonetti et al., 2009)). The initiation complex then associates with the 50S ribosomal subunit to form a translationally active 70S ribosome. The RBS comprises the start codon and Shine-Dalgarno sequence (SD) located somewhat upstream. MicF, the first trans-encoded sRNA characterized, base-pairs with the RBS of the <i>ompF</i> mRNA, thereby blocking its translation (Schmidt et al., 1995; Delihas and Forst, 2001). Binding of a sRNA within the approximately 55 nt region contacted by the ribosome (Hüttenhofer and Noller, 1994) can sterically hinder 30S ribosomal subunits from interacting with the RBS, thus preventing translation (Figure 4A). This region includes the first 5 codons of the open reading frame (Bouvier et al., 2008). This mode of regulation has subsequently been established for many more sRNA-mRNA pairs, e.g. MicA/<i>ompA</i> (Udekwu et al., 2005) OxyS/<i>fhlA</i> (Altuvia et al., 1998; Argaman and Altuvia, 2000), SgrS/<i>ptsG</i> (Kawamoto et al., 2006; Maki et al., 2008) OmrA and OmrB/<i>ydaM</i> (paper I), and is probably the dominant mode of action, at least in enterobacteria. When sRNAs block translation, the absence of ribosomes usually renders the mRNA susceptible to degradation by RNases. Therefore, most translationally inhibited mRNAs are also rapidly degraded (Wagner and Romby, 2015).

There are also examples where sRNAs bind further away from the RBS and still interfere with translation initiation (Figure 4A). Efficient translation initiation requires a relatively unstructured RBS. Yet, some mRNAs have a
structured RBS and are translated efficiently regardless. Several solutions to this problem are known, such as ribosome standby sites and translational coupling (reviewed in (Unoson and Wagner, 2007)). The standby model proposes that the 30S ribosomal subunit nonspecifically binds an accessible region in the mRNA, the ribosome standby site, so that upon transient opening of the inhibitory RBS structure, the 30S can relocate to form a canonical initiation complex (de Smit and van Duin, 2003). A ribosome standby site can be targeted by sRNAs to inhibit translation, as has been shown for the IstR-1 sRNA which binds a unstructured region about 100 nt upstream of the tisB RBS thereby blocking translation (Darfeuille et al., 2007; Wagner and Unoson, 2012a). Translational coupling occurs when the translation of one reading frame in an operon mRNA depends on the translation of the immediately preceding one. This is often a mechanism to overcome inhibitory structures. After translating the upstream ORF, ribosomes can proceed through the structure of the next RBS, and if stop and start codons are appropriately spaced, continue translation of the downstream ORF without dissociating from the mRNA. An example of this is translation of repA which is dependent on translation of a short upstream reading frame, tap. By binding immediately upstream of the tap SD, the cis-encoded CopA antisense RNA inhibits repA translation (Blomberg et al., 1992; Blomberg et al., 1994).

sRNAs may also bind to translational enhancers. E.g., GcvB downregulates several mRNA targets by binding to C/A-rich enhancer elements far upstream of the RBS (Sharma et al., 2007). It was recently suggested that OmrA and OmrB repress FepA translation by targeting an activating stemloop structure downstream of the RBS (Jagodnik et al., 2017). OmrA and OmrB inhibit CsgD translation by binding upstream, outside of the region contacted by the 30S ribosomal subunit, by a so far unknown mechanism (Holmqvist et al., 2010).

There are reports of sRNAs binding upstream of an RBS, resulting in translational inhibition, where Hfq rather than the bound sRNA causes repression. In these cases, the sRNA functions to recruit or stabilize Hfq on the messenger in close proximity to the RBS. (Desnoyers and Masse, 2012; Azam and Vanderpool, 2017)

**Inhibition by sRNA-mediated mRNA decay**

When sRNAs block translation, the absence of ribosomes renders the mRNA susceptible to degradation by RNases. Therefore, most translationally inhibited mRNAs are also rapidly degraded (Wagner and Romby, 2015). The formation of a ternary complex between the endonuclease RNase E, Hfq and sRNA could further explain how sRNAs mediate fast destabilization of target mRNAs (Morita et al., 2005). Nevertheless, it is suspected that, in most cases, translational inhibition is the principal regulatory event, followed by degradation, though only few cases have been studied in detail (Aiba, 2007; Wagner and Romby, 2015).
sRNAs can also directly mark mRNAs for degradation by RNase E, without effecting translation initiation (Figure 4B). The MicC sRNA base-pairs with *ompD* mRNA far into the coding region, and induces RNase E cleavage downstream of the duplex (Pfeiffer *et al.*, 2009). It was shown for MicC-*ompD* that duplex formation actively stimulates RNase E activity via a 5’ monophosphate on the sRNA interacting with the 5’ sensing pocket of RNase E, stimulating cleavage (Bandyra *et al.*, 2012). Since most sRNAs are primary transcripts and thus have a 5’ triphosphate, this would require the sRNA to be processed by a pyrophosphorhydrolase, or be matured by RNase E cleavage from a primary transcript (Bandyra *et al.*, 2012). Other examples of sRNAs targeting *ompD* mRNA far into the coding region inducing mRNA decay are SdsR, RybB, and likely also InvR (Pfeiffer *et al.*, 2007; Papenfort *et al.*, 2010; Fröhlich *et al.*, 2012). sRNAs are sometimes co-degraded with their mRNA targets (Massé *et al.*, 2003).

**Figure 4.** Schematic representation of regulatory mechanisms deployed by sRNAs to inhibit gene expression. (**A**) Translational inhibition by direct competition with initiating ribosomes. And variations; prevent 30S binding by blocking a ribosome standby site, and inhibition of translational coupling by binding the RBS of the upstream ORF. (**B**) mRNA degradation by active recruitment of RNase E.
sRNA-mediated target activation

As outlined above sRNAs are most commonly known to inhibit target gene expression. However, sRNAs can also positively affect mRNA translation and stability through various mechanisms (reviewed in (Fröhlich and Vogel, 2009; Papenfort and Vanderpool, 2015)) (Figure 5). Some mRNAs are translated inefficiently if at all, because their RBS is sequestered in a stable stemloop structure. Binding of an sRNA to one side of the stemloop breaks the structure, freeing the RBS, thus enhancing translation. This type of target activation is called anti-antisense (Figure 5A). It was first discovered for RNAIII-mediated activation of hla in the Gram-positive bacterium Staphylococcus aureus (Morfeldt et al., 1995). One of the best studied cases in Gram-negative bacteria is the mRNA encoding the stress Sigma factor RpoS, that is translationally activated by three sRNAs: DsrA (Lease et al., 1998; Majdalani et al., 1998), RprA (Majdalani et al., 2002), and ArcZ (Mandin and Gottesman, 2010). Another example of anti-antisense target mRNA activation in enterobacteria is that of the shiA mRNA by RyhB (Prévost et al., 2007).

sRNAs can also upregulate translation by stabilizing the mRNA. Some sRNAs bind to, and thereby mask, RNase recognition sites in mRNAs, thus inhibiting decay (Figure 5B). For example, RydC base-pairs with a RNase E recognition sequence in the 5’ UTR of cfa (Fröhlich et al., 2013). SgrS binds towards the 3’ end of the pldB coding sequence, sequestering a RNase E cleavage site, stabilizing the pldB-yigL mRNA after (SgrS-independent) processing and enhancing YigL translation (Papenfort et al., 2013). Stabilization can also be achieved through sRNA-guided processing of a mRNA into a more stable transcript; e.g. the cis-encoded GadY sRNA promotes processing of the gadX-gadW mRNA into separate more stable transcripts (Opdyke et al., 2004; Opdyke et al., 2011).

Traps and sponges

RNAs can indirectly affect expression of sRNA target genes by base-pairing with the sRNA, thereby sequestering it and often inducing decay. Such RNAs, which can be mRNAs or other sRNAs, have been called traps, decoys, sponges, and competing endogenous RNAs. Examples include the polycistronic chb mRNA. Upon pairing of ChiX to an intercistronic binding site, the sRNA is degraded, resulting in de-repression of chiP, its major target (Overgaard et al., 2009; Figueroa-Bossi et al., 2009). The SroC sRNA, which is generated by RNase E-dependent processing of the 3’ portion of the gltI mRNA, sequesters GcvB and promotes its degradation (Miyakoshi, Chao, and Vogel, 2015b).
Figure 5. Schematic representation of regulatory mechanisms deployed by sRNAs to activate gene expression. (A) anti-antisense (B) mRNA stabilization.

**Protein sequestration**

Some sRNAs affect gene expression by specifically binding and sequestering regulatory proteins, rather than via canonical antisense mechanisms. Probably the most well-known example is the CsrB/RsmZ family of sRNAs. In *E. coli*, CsrB and CsrC sequester CsrA (carbon storage regulator) and titrate it away from its target mRNAs (reviewed in (Babitzke and Romeo, 2007; Romeo et al., 2013)). The RNA binding protein CsrA recognizes GGA motives in mRNAs to mostly affect translation and/or stability negatively, but sometimes positively. The sRNAs CsrB and CsrC contain multiple CsrA recognition sequences and thereby compete with mRNAs for CsrA binding.

Another example is the ubiquitous and abundant 6S RNA (reviewed in (Wassarman, 2007; Steuten et al., 2014). In *E. coli* the 6S RNA represses expression from σ^70 dependent promoters in stationary phase by directly binding to the σ^70 containing RNA polymerase (Wassarman and Storz, 2000). Binding relies on the conserved secondary structure of 6S, which resembles an open promoter (Wassarman and Saecker, 2006).

**The RNA chaperone Hfq**

Hfq was first discovered as an essential host factor for the replication of phage Qβ, hence the name Hfq (host factor Qβ) (Franze de Fernandez et al., 1968). Hfq belongs to the Sm/LSm protein family. These are RNA binding multi-meric ring-like protein complexes that are present in all kingdoms of life (Mura et al., 2013). Hfq associates with diverse RNAs but primarily with mRNAs and sRNAs (Zhang et al., 2003; Sittka et al., 2008; Chao et al., 2012; Holmqvist et al., 2016). Hfq stabilizes RNAs and facilitates base-pairing of trans-encoded sRNAs and their mRNA targets (Vogel and Luisi, 2011; Uptegrove et al., 2016). Stabilization of RNAs by Hfq is probably due to the
similarity between Hfq-binding and RNase E recognition sites (Moll et al., 2003). As Hfq is required for this large network of post-transcriptional regulators, it is not surprising that, in enterobacteria, deletion of Hfq reduces fitness, impairs virulence and stress responses (Chao and Vogel, 2010; Sobrero and Valverde, 2012).

Figure 6. Cartoon representing the crystal structure of Hfq (C-terminally truncated after S72) (PDB-ID 2YLB). (A) Single monomer showing the LSm motive. (B) Top and side view of the Hfq hexameric ring are shown. A single monomer is highlighted in grey, and RNA binding surfaces are indicated. Molecular graphics were generated using the UCSF Chimera package (Pettersen et al. 2004).

Structure and RNA binding properties
While the classical eukaryotic Sm and LSm proteins form heteroheptamers, Hfq assembles into a homohexameric ring (Figure 6) (Sauer, 2013). Each monomer comprises a Sm domain, which consists of a N-terminal α-helix followed by five β-strands separated by loops (Figure 6A), and an intrinsically disordered C-terminal domain (Sauer, 2013). The site on which the N-terminal α-helix is exposed is called the proximal face and binds preferentially U-rich sequences (Schumacher et al., 2002), such as those found in Rho-independent transcription terminators (Otaka et al., 2011; Ishikawa et al., 2012; Holmqvist et al., 2016). Specific recognition of the 3’ hydroxyl group increases affinity and might explain how Hfq protects RNAs from 3’ exonucleases (Sauer and Weichenrieder, 2011). While the 3’ end of the sRNA is anchored at the proximal side, internal A/U rich sequences in the sRNA body are bound by arginine patches on the lateral rim of Hfq (Sauer et al., 2012). Opposite to the proximal face is the distal face. It recognizes A-rich sequences (Link et al., 2009), or more specifically AAN-motifs in E. coli (N = any nucleotide) (Robinson et al., 2014), and interacts primarily with mRNAs. The majority of sRNAs interact with the proximal and rim face of Hfq, and mRNAs with the distal face (Schu et al., 2015). However, some sRNAs also interact with the distal face, while some mRNAs require rim residues for binding (Schu et al., 2015). So, there is likely a whole spectrum of RNA-Hfq interactions, depending on the nature and positioning of Hfq binding motifs in the RNA.
The location of the Hfq recognition motifs within the RNAs is also of consequence. Positioning of the AAN-motif close to, but not overlapping with, the sRNA base-pairing region on the mRNA, enhances sRNA-mediated annealing (Beisel et al., 2012; Panja and Woodson, 2012; Peng, et al., 2014). A global analysis of Hfq binding sites on RNAs in *Salmonella enterica* using crosslinking immunoprecipitation followed by high-throughput sequencing (CLIP-seq) revealed that Hfq is significantly more likely to bind on the 5’ side of the sRNA target site in an mRNA, but more often towards the 3’ of the sRNA seed sequence. This should facilitate simultaneous access of sRNA and mRNA target to Hfq and position complementary sequences for pairing (Holmqvist et al., 2016).

The sequence of the C-terminal domain (CTD) is not conserved, and its length varies widely between bacterial families (Sun et al., 2002; Sobrero and Valverde, 2012). Its role in sRNA-mediated regulation has been somewhat controversial due to conflicting results. Some reports showed that deletion of (part of) the CTD did not affect sRNA-mediated regulation. Others suggested a dependency for duplex formation (reviewed in (Sobrero and Valverde, 2012)). A recent study demonstrated that the CTD does not stimulate duplex formation, but rather aids in the release of formed RNA complexes (Santiago-Frangos et al., 2016). This is important for Hfq turnover and has implications for the competition between sRNAs for Hfq access (see below ‘cycling of RNAs on Hfq’) (Santiago-Frangos et al., 2016). Displacement of double stranded RNA is mediated by short-lived interactions between the acidic tip of the CTD and the basic arginine residues on the Hfq rim (Santiago-Frangos et al., 2017).

**Promotion of sRNA-mRNA pairing**

As discussed above, sRNAs and their mRNA targets can simultaneously bind to a Hfq hexamer, resulting in sRNA-mRNA duplex formation. It is generally considered that there are two, not mutually exclusive, ways by which Hfq can promote RNA annealing (Vogel and Luisi, 2011; Sobrero and Valverde, 2012). By binding to both RNAs simultaneously, Hfq may simply bring them into close proximity to aid duplex formation. Hfq can also induce structural rearrangements in RNAs. This has been shown to actively position RNAs for optimal base-pairing. Again, a good example is the *rpoS* mRNA. By contacting the distal and rim surfaces, the mRNA is folded into a compact ternary structure that positions the base-pairing region close to the DsrA seed-sequence upon binding (Peng, et al., 2014; Peng, et al., 2014). sRNAs can also be restructured on Hfq to favor duplex formation, as exemplified by RydC and its target *cfa* (Dimastrogiovanni et al., 2014).

Both these mechanisms consider Hfq as a platform for RNA interaction, where both RNA molecules meet on Hfq. In paper I we describe a system in which Hfq serves as a chaperone, unfolding an inhibitory structure in the
mRNA target that prevents the sRNAs from binding, but with no requirement for concurrent binding of both RNAs to Hfq.

**Cycling of RNAs on Hfq**

sRNAs and their mRNA targets need to converge on the same Hfq molecule in order for Hfq to enhance duplex formation. Given the high total concentration of binding-competent RNAs present in the cell, Hfq is probably saturated at all times (Wagner, 2013). Indeed, it has been shown that sRNAs compete for Hfq *in vivo* (Hussein and Lim, 2011; Moon and Gottesman, 2011). So, for a newly induced sRNA to gain access to Hfq and perform its regulatory function, another RNA needs to come off. In other words, the pool of Hfq-bound RNAs needs to rapidly exchange in order to reflect any changes in gene expression (Wagner, 2013). As measured dissociation rates of RNA-Hfq complexes are slow, with half-lives in the order of hours, ((Fender *et al.*, 2010 and references therein), passive dissociation cannot account for the fast 1-2 minute *in vivo* response times of sRNAs (Massé *et al.*, 2003; Papenfort *et al.*, 2006). Instead, studies showed that RNAs actively cycle on Hfq, driven by the concentration of free competitor RNA (Salim and Feig, 2010; Fender *et al.*, 2010; Olejniczak, 2011). sRNAs differ in their ability to compete, which seems to depend on the sRNAs potential to interact with the different binding surfaces of Hfq (Fender *et al.*, 2010; Olejniczak, 2011; Malecka *et al.*, 2015). A recent report shows that the Hfq C-terminal domain might contribute to this hierarchy among competing sRNAs (Santiago-Frangos *et al.*, 2016).

**Other RNA binding proteins**

Not all bacteria encode Hfq, which begs the question whether they contain another type of RNA chaperone that fulfills this function. Additionally, some sRNAs do not require Hfq for base-pairing. These sRNAs may make use of an alternative RNA binding protein (RBP). Recently, one such alternative RBP, containing a FinO domain, has received much attention (reviewed in (Attaiech *et al.*, 2017; Olejniczak and Storz, 2017)). In *Salmonella*, ProQ, a FinO domain-containing protein, associates with a wide range of RNAs, including many sRNAs (Smirnov *et al.*, 2016). These were predominantly cis-encoded sRNAs, but also included trans-encoded sRNAs, some of which are also known to interact with Hfq (Smirnov *et al.*, 2016). Like Hfq, ProQ stabilizes RNAs that it associates with (Smirnov *et al.*, 2016). It was further shown that ProQ aids translational repression of *hupA* by the *Salmonella* sRNA RaiZ (Smirnov *et al.*, 2017).

**Advantages of regulating with sRNAs**

As mentioned earlier, sRNAs can be found in virtually any bacterial species, and are part of most global regulatory networks. Why are sRNAs so omnipresent? The answer probably lies in the unique features of sRNAs, which
complement regulation by TFs (Beisel and Storz, 2010; Wagner and Romby, 2015; Nitzan et al., 2017).

sRNAs can establish a threshold for gene expression. If the rate of sRNA production is higher than that of the target mRNA, all mRNA will promptly be bound by sRNA and expression is silenced. However, once production of the mRNA exceeds this threshold, expression increases linearly (Levine et al., 2007; Levine and Hwa, 2008). This phenomenon, called threshold-linear response, might function to filter out transcriptional noise (reviewed in (Levine and Hwa, 2008)). Due to its stochastic nature, rare bursts of transcription will occur, even from repressed genes. These bursts of mRNAs could generate an even larger burst of proteins, because multiple rounds of translation can occur on one mRNA. sRNAs can counteract this noise by preventing translation and/or inducing destabilization of these stochastically arising mRNAs, thereby maintaining a silenced state. This could be especially important when a large burst of proteins might affect the phenotype of the cells, for example by producing the biofilm matrix component curli, or SOS-induced toxin expression (Wagner and Romby, 2015; Holmqvist and Wagner, 2017). Conversely, the threshold-linear model predicts ultra-sensitivity near the threshold, and could therefore also generate stochastic fluctuations leading to increased phenotypic diversity (Beisel and Storz, 2010; Wagner and Romby, 2015). Similarly, sRNAs can filter out transient signals, while still allowing for a rapid response when the signal persists (Levine and Hwa, 2008). As the threshold is set by the stoichiometry between sRNA and mRNA, this can lead to cross-talk between multiple targets of the same sRNA, opening up the possibility of a hierarchy of responses based on an mRNA's binding affinity (Levine and Hwa, 2008).

Many sRNAs are induced upon various environmental stresses which require rapid changes in gene expression. A plausible explanation for the enrichment of sRNAs in these types of networks is their faster regulatory speed. When a TF represses transcription, the already available mRNAs keep being translated until they are naturally turned over. Because sRNAs act post-transcriptionally, they can directly inhibit further protein synthesis and destabilize mRNAs (Shimoni et al., 2007; Nitzan et al., 2017). The faster response provided by sRNAs can be vital, for example in the response to outer membrane stress (Papenfort et al., 2006; Holmqvist and Wagner, 2017). When misfolded proteins accumulate in the outer membrane, σ^E is induced and activates the expression of proteases and chaperones to degrade and refold misfolded proteins. To provide fast stress relief by preventing further expression of harmful outer membrane proteins, σ^E induces expression of three sRNAs: MicA, RybB and MicL ((Holmqvist and Wagner, 2017) and references therein). The importance of these sRNAs in preventing membrane stress is highlighted by the observation that, in their absence, the membrane stress response is induced, and that overexpression can rescue growth arrest in σ^E-deficient cells (Papenfort et al., 2006; Gogol et al., 2011; Guo et al., 2014).
The OmrA and OmrB sRNAs

The *E. coli* sRNAs OmrA and OmrB are 88 and 82 nt long, respectively (Argaman *et al.*, 2001; Wassarman *et al.*, 2001; Vogel *et al.*, 2003). They likely arose through a gene duplication event, and are well conserved within enterobacterial species (Skippington and Ragan, 2012). OmrA and OmrB associate with Hfq (Wassarman *et al.*, 2001; Zhang *et al.*, 2003; Holmqvist *et al.*, 2016), and this is required for their regulatory activity and stability (Guillier and Gottesman, 2008; Holmqvist *et al.*, 2010).

Omr stands for OmpR regulated sRNA, as transcription of OmrA and OmrB is induced by the OmpR-EnvZ two component system upon osmotic and other surface stress (Guillier and Gottesman, 2006; Brosse *et al.*, 2016). The Omr sRNAs show a somewhat different expression pattern, with OmrA accumulating towards stationary phase, while OmrB is present throughout logarithmic as well as stationary phase ((Argaman *et al.*, 2001; Wassarman *et al.*, 2001)). This could be explained by transcription of OmrA, but not OmrB, responding to Sigma factor $\sigma^s$ (Lévi-Meyrueis *et al.*, 2014; Peano *et al.*, 2015; Colgan *et al.*, 2016). Dual RNA-seq showed induction of OmrA and OmrB upon *Salmonella enterica* infection of HeLa cells (Westermann *et al.*, 2016); this induction could occur through OmpR-EnvZ or (in combination with) a different TF.

OmrA and OmrB have nearly identical 5’ and 3’ tails, but differ in their central region (Figure 7). Their 5’ regions are used for base-pairing with target mRNAs, and thereby downregulating their overlapping regulon. Known targets are the outer membrane proteins CirA, FecA, and FepA which are involved in iron uptake, and the outer membrane protease OmpT (Guillier and Gottesman, 2006; Guillier and Gottesman, 2008; Jagodnik *et al.*, 2017). The sRNAs also form a negative feedback loop with their own activator OmpR (Guillier and Gottesman, 2008; Brosse *et al.*, 2016). Furthermore, OmrA and OmrB regulate several targets in the inversely controlled pathways for motility and biofilm formation. They inhibit the principal regulator of flagella synthesis, *flhD* (De Lay and Gottesman, 2012), as well the anti-Sigma factor *flgM* (Hoekzema *et al.* in preparation). While inhibition of *flhD* prevents cells to take the first steps towards flagella synthesis, the inhibition of *flgM* has a positive effect on expression of late flagella genes. This somewhat puzzling con-

Figure 7. Figure indicating the sequence similarities between OmrA and OmrB. An . indicates that the nucleotide is the same in both sRNAs, an – indicates a gap in OmrB. The 5’-end sequence used for base-pairing with known targets is encircled.
trast needs further investigation. The role of OmrA and OmrB to impede biofilm formation is evident by inhibition of both the master regulator for biofilm formation \(csgD\), as well as the DGC \(ydaM\) which is required for \(csgD\) transcription ((Holmqvist et al., 2010), paper I).

sRNAs in motility and biofilm formation

sRNA-based regulation is an important part of many global regulatory networks involving metabolism, stress responses, and pathogenesis (Michaux et al., 2014; Wagner and Romby, 2015). Since I cannot cover all, I will only provide some details on the many sRNAs involved in control of motility and biofilm formation, as this is relevant to paper I included in this thesis.

In the transition from active division in exponential, to starvation in stationary phase, cells go through a successive expression of flagella followed by biofilm matrix components. These inversely regulated processes are at the transcriptional level controlled by Sigma factor competition, a hierarchical cascade of TFs, as well as second messengers. In recent years, it has become clear that, at the post-transcriptional level, a great number of sRNAs contribute to the decision making by regulating key targets in both the flagella and biofilm regulatory cascades, thereby fine tuning their expression and integrating additional environmental signals into the network (reviewed in (Mika and Hengge, 2013; Mika and Hengge, 2014)) (Figure 8).

Flagella are whip-like appendages on the cell surface that function to propel bacteria forward in liquid and on surfaces. The *E. coli* flagella operons are organized in three classes that are temporally regulated with regard to flagella assembly (reviewed in (Chevance and Hughes, 2008)). On top of the cascade, making the decision whether to produce flagella, is \(flhDC\). The FlhD4C2 complex is required for \(\sigma^{70}\)-dependent activation of the Class 2 flagellar genes that are necessary for the structure and assembly of the flagellar motor, the hook-basal body. The class 2 genes also contain \(fliA (\sigma^{28})\) and \(flgM\) (anti-\(\sigma^{28}\)). Upon completion of the hook-basal body, FlgM is secreted, and \(\sigma^{28}\)-containing RNA polymerase can transcribe the Class 3 genes that code for the filament structural components, as well as chemotaxis proteins. In addition to OmrA and OmrB, four more sRNAs, ArcZ, OxyS, SdsR, and GadY, negatively regulate \(flhDC\) expression (De Lay and Gottesman, 2012). Additionally, the McaS sRNA positively affects FlhD translation (Thomason et al., 2012; De Lay and Gottesman, 2012).
Figure 8. Simplified representation of the regulatory network controlling flagella and biofilm component expression, showing its complex interconnected nature. Grey lines: transcriptional regulation, black lines: post-transcriptional regulation, dotted lines: direct protein-protein or protein-c-di-GMP interactions. Activating interactions: arrows, repressing interactions: blunt ended lines.

At ambient temperature, the *E. coli* biofilm matrix is composed of amyloid curli fibers, cellulose, and colonic acid, while flagella are used for initial attachment (reviewed in (Beloin *et al.* 2008)). The Sigma factor $\sigma^s$ is on top of the biofilm regulatory cascade, followed by the TF MlrA. MlrA activity is regulated by a c-di-GMP-dependent switching module consisting of the phosphodiesterase (PDE) YciR and the diguanylate cyclase (DGC) YdaM. YciR inhibits YdaM until c-di-GMP levels are sufficient for it to start functioning as a PDE. Then, the de-repressed YdaM activates MlrA, and produces more c-di-GMP, creating a positive feedback loop (Lindenberg *et al.* 2013). Both MlrA and $\sigma^s$ are required for transcription of the master regulator of biofilm formation *csgD*, which in turn activates transcription of the *csgBAC* operon coding for the curli structural genes, and *yaiC*, encoding a DGC essential for cellulose production. sRNAs regulate on all three levels of the biofilm regulatory cascade. The $\text{rpoS}$ SD sequence is concealed in an inhibitory structure and therefore inefficiently translated. Base-pairing of the sRNAs DsrA, RprA, or ArcZ, upstream of the SD restructures the mRNA so that the RBS becomes accessible for translation (Lease *et al.*, 1998; Majdalani *et al.*, 1998; Majdalani *et al.*, 2002; Mandin and Gottesman, 2010). The next regulatory hub for signal integration in biofilm formation is the *csgD* mRNA (Boehm and Vogel, 2012). Besides OmrA and OmrB (Holmqvist *et al.*, 2010), four more sRNAs, McaS, RprA, GcvB, and RydC, negatively regulate *csgD* expression (Jørgensen *et al.*, 2012; Thomason *et al.*, 2012; Mika *et al.*, 2012; Bordeau and Felden,
RprA and OmrA/B additionally indirectly inhibit CsgD production by down-regulating ydaM expression ((Mika et al., 2012; Mika and Hengge, 2013), paper I).

Toxin-antitoxin systems

Toxin-antitoxin (TA) systems consist of two components: a stable toxin that either kills the cell or causes growth arrest, and an unstable antitoxin that neutralizes the toxin. Currently, six types of TA systems are known, which differ in the nature of the antitoxin as well as mechanisms used to counteract the toxin (Page and Peti, 2016). Types II, IV, V and VI have an antitoxin protein, for type I and III the antitoxin is a RNA. The toxin is in all cases a protein. In the case of type I TA systems, the antitoxin sRNA acts by base-pairing to the toxin mRNA. By contrast, in type III TA systems, the antitoxin RNA directly blocks the toxin's activity via a pseudoknot structure. For type II TA-systems, the antitoxin protein directly inhibits the toxin protein, while in type IV, the antitoxin interferes with target binding. Type V antitoxins cleave the toxin mRNA, and type VI antitoxins promote degradation of the toxin itself. TA loci are found on plasmids as well as the chromosomes of free-living bacteria ((Pandey and Gerdes, 2005; Gerdes and Wagner, 2007) and references therein). Three biological functions have been described for TA systems, post-segregational killing, abortive infection, and persister cell formation (Harms et al., 2018). Post-segregational killing relies on the difference in stability of the toxin and antitoxin to stabilize mobile elements (Harms et al., 2018). After cell division, a plasmid-free daughter cell experiences rapid loss of the unstable antitoxin, resulting in cell death by remaining, stable toxin. Post-segregational killing can similarly stabilize chromosomal regions. Abortive infection kills phage infected cells prior to phage replication, and thereby impedes phage propagation in a bacterial population (reviewed in Dy et al., 2014). Persisters are a subpopulation of dormant cells that are transiently tolerant to antibiotics and other stress conditions. The role of TA-systems in persister cell formation is discussed below and reviewed in (Gerdes and Maisonneuve, 2012; Berghoff and Wagner, 2017; Harms et al., 2018).

Type I toxin-antitoxin systems

The sRNA antitoxins of type I TA systems can be cis- or trans-encoded. However, in contrast to what was discussed for other trans-encoded sRNAs, apparently antitoxin sRNAs only target the toxin mRNA and are largely Hfq independent (Brantl, 2012; Brantl and Jahn, 2015). Interestingly, the IstR-1 antitoxin interacts with ProQ, opening up for the compelling possibility that this protein is involved in mediating type I toxin-antitoxin interactions.
((Smirnov et al., 2016), and Erik Holmqvist personal communications). Anti-toxin sRNAs inhibit their mRNA targets in much the same way as sRNAs in general, either by translational inhibition or by active degradation (see ‘sRNA modes of action’). Besides antitoxin-mediated regulation, additional layers of control ensure tight regulation of toxin expression. Many toxin mRNAs are highly structured and require processing before the mRNA can be translated, and other examples include rare start codons and SD sequences that slowly release ribosomes (Brantl, 2012; Berghoff and Wagner, 2017). Almost all type I toxins are small hydrophobic proteins. Some of these insert into the inner membrane leading to depolarization and ATP depletion (Brantl and Jahn, 2015; Berghoff and Wagner, 2017).

The \( \text{tisB/istR-1} \) toxin-antitoxin system

The \( \text{tisB/istR1} \) locus consists of \( \text{tisB} \) which is under transcriptional control of LexA, the master regulator of the SOS response, an upstream ORF \( \text{tisA} \) which is not translated nor required for toxicity, and two sRNAs, \( \text{IstR-1} \) and \( \text{IstR-2} \), transcribed divergently in the opposite direction (Figure 9) (Vogel et al., 2004; Darfeuille et al., 2007). The 74 nt long \( \text{IstR-1} \) sRNA is constitutively expressed and translationally inhibits \( \text{tisB} \) mRNA. \( \text{IstR-2} \) (140nt) is, like \( \text{tisB} \), under LexA control and is not involved in \( \text{tisB} \) regulation. Its function remains unknown (Vogel et al., 2004; Darfeuille et al., 2007).

The primary \( \text{tisB} \) transcript, denoted +1 (number referring to 5’ end position) is translationally inactive because an inhibitory structure blocks the RBS. Processing of the +1 transcript generates the translation-competent +42 mRNA in which a ribosome standby site has become accessible. It is this standby site that is targeted by \( \text{IstR-1} \) to indirectly inhibit expression of \( \text{tisB} \). Co-degradation of the \( \text{tisB-IstR-1} \) complex by RNase III generates a second translationally inactive, +106, \( \text{tisB} \) mRNA. (Darfeuille et al., 2007)

So, under normal growth conditions, \( \text{tisB} \) transcription is repressed by LexA, and any escaping \( \text{tisB} \) +42 transcripts in spite of transcriptional repression are neutralized by \( \text{IstR-1} \). Upon SOS induction, when \( \text{tisB} \) transcription is on, the high mRNA levels out-titrate \( \text{IstR-1} \), eventually leading to \( \text{TisB} \) toxin production (Wagner and Unoson, 2012b).

Role in persister formation

Bacterial persisters are a small subpopulation of dormant cells that are transiently antibiotic-tolerant, since most antibiotics only kill actively growing/ dividing cells (Bigger, 1944; Lewis, 2010) (Figure 10). Within a normally growing, genetically identical population, some cells will stochastically enter a persister state (Balaban et al., 2004). This generation of phenotypic heterogeneity is referred to as bet-hedging, and is a common survival strategy in bacteria (Veening et al., 2008). In some cases, increasing numbers of persisters can be
induced in response to different types of stress (Harms et al., 2016). This is called responsive diversification (Kotte et al., 2014).

Activation of TA toxins can serve to induce transition into a dormant state by inhibition of vital cellular processes. In E. coli, multiple TA systems have been implied in persister formation (reviewed in (Gerdes and Maisonneuve, 2012; Berghoff and Wagner, 2017; Harms et al., 2018)). Among these is the type I tisB/IstR-1 TA system discussed above. The small hydrophobic TisB toxin inserts itself into the inner membrane forming pores, resulting in loss of membrane potential, a drop in intracellular ATP concentration, and growth arrest ((Unoson and Wagner, 2008; Gurnev et al., 2012), paper II). In this way, the tisB/istR-1 TA system could impact persister cell formation. Congruent with this, after induction of the SOS response, persister levels increase in a DistR-1 strain, while they decrease in a DistB background ((Dörr et al., 2010), paper II). It should be noted that besides TA modules other factors have been proposed to be involved in persister cell formation (Harms et al., 2016).

The microbial scout model hypothesizes cells to leave the dormant state in a stochastic manner. If conditions are favorable (the antibiotic is removed), these cells then repopulate the environment (Buerger et al., 2012). However, the molecular mechanisms behind persister resuscitation remain largely unknown. The newly emerging population is as sensitive to antibiotics as the original one, since there are no genetic differences, and will produce a similar fraction of persisters (Figure 10). Persister cells can cause relapsing bacterial infections and are a public health concern (Lewis, 2010; Fisher et al., 2017).
It should be mentioned that there is a considerable controversy in the field on the role of TA systems in persister cell formation. A recent paper demonstrated that a strain deleted for ten type II TA systems, which previously was reported to exhibit reduced persister cell formation, had in fact been infected by a lysogenic phage that caused the phenotype (Harms et al., 2017). The same study also addressed inconsistencies that readily arise from small variations in protocols commonly used to study persister formation, and suggests a new more controlled approach (Harms et al., 2017).

Interestingly, it has recently been reported that, in a fashion similar to TA-systems, also the toxin and immunity proteins of the *E. coli* contact-dependent growth inhibition (CDI) systems can be involved in persister cell formation (Ghosh et al. in press).

Figure 10. Schematic of the killing curve of a bacterial population with a bactericidal antibiotic. After addition of the antibiotic at time zero the sensitive cells (black) rapidly die (black dotted line), until only the persister cells (grey) remain. Due to stochastic resuscitation events the persisters are killed but at a much slower rate (grey dotted line). Once the antibiotic disappears the remaining persister cells can repopulate the environment, this new population will be as susceptible to antibiotic treatment as the original. (Figure adapted from {Harms:2016de}.)
Figure 11. Schematic overview of the stages of CRISPR-Cas immunity. Adaptation is the integration of a new spacer into the CRISPR-array. During the expression stage the cas genes are transcribed and translated, and the pre-crRNA primary transcript of the CRISPR-array is processed into mature crRNAs. Together the crRNA and Cas proteins form a ribonucleoprotein effector complex. Interference is when the effector complex, guided by the crRNA, identifies a cognate target and proceeds to degrade the nucleic acid.

The CRISPR-Cas adaptive immune system

CRISPR-Cas is a prokaryotic adaptive immune system that defends against foreign genetic elements such as plasmids and phages. This RNA-mediated defense system targets and degrades foreign nucleic acids in a sequence specific manner. CRISPR-Cas loci are widespread in both bacteria and archaea (Makarova et al., 2015). Memory resides in the clustered regularly interspaced short palindromic repeats (CRISPR) array, where small snippets of previously encountered foreign DNA are inserted in the hosts chromosome. The CRISPR-array, first observed in E. coli in 1987, consists of a series of repeats alternated with unique spacer sequences of similar length (Ishino et al., 1987).
A breakthrough came with the observation that spacer sequences were primarily derived from foreign genetic elements, which led to the realization that this could be an immune system (Mojica et al., 2005; Pourcel et al., 2005; Bolotin et al., 2005). This was confirmed by a seminal study showing the CRISPR-Cas locus conferred phage resistance to *Streptococcus thermophilus* (Barrangou et al., 2007). The often adjacent and highly diverse CRISPR-associated (cas) genes encode the machinery driving immunity (Brouns et al., 2008). CRISPR-Cas systems are divided into two classes with several types and subtypes based on specific ‘signature’ cas genes (Makarova et al., 2015; Shmakov et al., 2015) (see below ‘Diversity and classification’). CRISPR-Cas immunity can be divided in three mechanistic stages (Figure 11): Adaptation, the incorporation of a new spacer in the CRISPR-array. Expression, the transcription of the CRISPR-Cas components and processing of the primary transcript of the CRISPR-array into mature CRISPR RNAs (crRNAs). And finally, interference, the crRNAs guide Cas endonucleases to the target nucleic acid leading to degradation. What is known about the mechanistic details of these three stages of CRISPR-Cas immunity will be discussed below.

Prokaryotes evolve strategies to counteract phage infection, and phages in turn evolve counter measures to those strategies. CRISPR-Cas immunity is no exception and by now several anti-CRISPR proteins have been discovered (reviewed in (Pawluk et al., 2018)). Interestingly, non-immune functions of CRISPR-Cas loci have been discovered as well (reviewed in (Westra et al., 2014)).

Diversity and classification

CRISPR-Cas systems are prevalent, occurring in ~50% of bacterial, and ~87% of archaeal chromosomes sequenced so far (Makarova et al., 2015). However, a recent study that used a cultivation independent approach found CRISPR-Cas systems in only 10% of 1724 sampled microorganisms, indicating that the presumed high prevalence of CRISPR-Cas systems might be due to a sampling bias (Burstein et al., 2016). In line with their role as immune system CRISPR-Cas evolution is dynamic and modular (Makarova et al., 2015). CRISPR-Cas systems are divided into two classes each with three types and several subtypes based on signature genes as well as overall architecture of the cas loci (Makarova et al., 2015; Shmakov et al., 2015). Class 1 systems are defined by a multiprotein crRNA effector complex, opposed to a single protein effector module in class 2 systems. Class 1 comprises type I subtypes A- F and U, type III subtypes A-D, and the putative type IV. Class 2 encompasses type II subtypes A-C, type V subtypes A-C and type VI. Not only the cas genes, but also the CRISPR-arrays are very diverse. They are classified into 18 structural and 24 sequence families (Kunin et al., 2007; Lange et al.,
The CRISPR-array families show a preferential association with certain types and sub-types of cas loci (Makarova et al., 2015).

The CRISPR-Cas related works included in this thesis have made use of the type I-E system of E. coli as a model (Figure 12). The hallmark of this system is the CRISPR-associated ribonucleoprotein complex for antiviral defense (Cascade), encoded by cse1, cse2, cas5e, cas6e, cas7. Cascade is responsible for processing of the pre-crRNA into mature crRNAs, as well as, in conjunction with the nuclease/helicase Cas3, interference (Brouns et al., 2008). The system further consists of Cas1 and Cas2 which mediate adaptation (Brouns et al., 2008; Yosef et al., 2012), and two CRISPR-arrays. CRISPR-I with 13 spacers and CRISPR-II with six spacers, where CRISPR-I is associated with the cas genes (Diez-Villaseñor et al., 2010). The spacers and palindromic repeats are 32 and 29 base-pair (bp) respectively.

![Figure 12. The E. coli CRISPR-Cas locus.](image)

**Adaptation**

CRISPR-Cas systems rely on the information stored in the CRISPR-array in the form of spacer to specifically recognize and neutralize mobile genetic elements. To keep in pace with the continuously changing pool of invaders the acquisition of new spacers is thus essential. The process of incorporating a new spacer into the CRISPR-array is termed adaption. The essential components for adaptation in the type I-E CRISPR-Cas system are Cas1, Cas2, part of the leader and one repeat sequence (Yosef et al., 2012; Diez-Villaseñor et al., 2013). Additionally, several host factors have been implicated in adaptation (Levy et al., 2015; Ivančić-Baće et al., 2015; Nuñez et al., 2016). It should be noted conflicting evidence exists regarding the requirement for recB (Levy et al., 2015; Ivančić-Baće et al., 2015). Two modes of adaptation, naïve and primed, have been reported for the E. coli type I-E system. Where in naïve adaptation no previous memory exists, in primed adaptation pre-existing spacers are utilized to acquire additional spacers targeting the same invader (Swarts et al., 2012; Datsenko et al., 2012). Primed adaptation consequently requires all CRISPR-Cas components. Spacers are preferentially integrated at the leader proximal end of the CRISPR-array, which optimizes the immune response against recently encountered threats (McGinn and Marraffini, 2016). The CRISPR-array therefore gives a chronological account of the hosts phage infections, which can be used to study phage-host evolution and ecology (e.g. (Andersson and Banfield, 2008)).
Figure 13. (A) Crystal structure of the Cas1-Cas2 complex bound to a prespacer. (PDB-ID 5DQZ). Cas1 and Cas2 dimers as well as the prespacer and PAM sensing pocket are indicated. Molecular graphics were generated using the UCSF Chimera package (Pettersen et al. 2004). (B) Step wise schematic of the integration process. After the Cas1-Cas2-prespacer complex is docked at the leader repeat boundary the first nucleophilic attack occurs at the leader-repeat junction resulting in a half-site intermediate. The second nucleophilic attack occurs at the repeat-spacer junction resulting in full site integration. Host repair enzymes fill in and ligate the repeats.
Spacer integration

The proteins primarily responsible for catalyzing spacer integration are Cas1 and Cas2. Despite the diversity of CRISPR-Cas systems discussed previously, Cas1 and Cas2 are relatively conserved, and associated with most types of CRISPR-Cas systems (Makarova et al., 2015). In the *E. coli* type I-E system Cas1 and Cas2 form a complex with two Cas1 dimers bridged by a central Cas2 dimer (Cas1-Cas2) (Nuñez et al., 2014; Wang et al., 2015; Nuñez et al., 2015) (Figure 13A). While Cas2 is essential, Cas2 nuclease activity is not required for integration (Nuñez et al., 2014). The preferred substrate for the Cas1-Cas2 complex is double stranded DNA (dsDNA) with 3’ overhangs (Arslan et al., 2014; Nuñez et al., 2015). In order for leader-polarized spacer integration to take place the Cas1-Cas2 complex loaded with the prespacer must correctly locate the leader repeat boundary. In the type I-E system this recognition is assisted by the integration host factor (IHF) binding to the leader and inducng DNA bending (Nuñez et al., 2016; Yoganand et al., 2017). Additionally, the palindromic repeat sequences could form a cruciform DNA structure that might be recognized by Cas1-Cas2 (Arslan et al., 2014; Nuñez et al., 2015). Spacer integration proceeds through two nucleophilic attacks, one on each strand of the leader proximal repeat, by the 3’ OH ends of the Cas1-Cas2 bound prespacer (Nuñez et al., 2015; Goren et al., 2016) (Figure 13B). This results in a fully integrated spacer flanked on either side by a single stranded repeat (Arslan et al., 2014). The repeats are subsequently filled in by host factors, likely DNA polymerase I (Ivančić-Baće et al., 2015) (Figure 13B).

The sequence in the target nucleic acids that is complementary to the spacer is called protospacer. To avoid self-targeting the type I-E CRISPR-Cas effector complex only recognizes and destroys DNA that has a specific short sequence next to the protospacer, called the protospacer adjacent motif (PAM) (see below ‘Interference’). In *E. coli* a majority of the spacers acquired by naïve adaptation target protospacers associated with a correct PAM, though exact numbers remain elusive due to ambiguity about the consensus PAM sequence (Yosef et al., 2012; Xue et al., 2015). A pocket in the Cas1 dimer recognizes the PAM complementary sequence in the prespacer prior to trimming (Figure 13A), and positions it appropriately relative to the cleavage site to ensure integration in the correct orientation (Wang et al., 2015).

Substrate capture

Acquisition of spacers from the host chromosome needs to be avoided, as such autoimmunity is usually detrimental (Stern et al., 2010). A strong preference for spacers derived from foreign DNA has been observed for the *E. coli* type I-E CRISPR-Cas system (Yosef et al., 2012; Diez-Villaseñor et al., 2013; Levy et al., 2015). This could not be explained by selection pressure against
chromosomal spacer acquisition, as the interference machinery was either absent or repressed. Therefore, there must be some inherent preference for non-self DNA. As over-expression of Cas1 and Cas2 leads to an increased fraction of chromosomal derived spacer, it appears Cas1 and Cas2 do not have an intrinsic ability to recognize exogenous DNA (Levy et al., 2015). So how is this preference for foreign DNA, and avoidance of self DNA created? It has been suggested that spacer substrates in *E. coli* are largely derived from RecBCD-mediated processing of double stranded DNA breaks, and that this is at least in part responsible for the self/non-self discrimination (Levy et al., 2015). After recognition of a double stranded break RecBCD unwinds and degrades the DNA, generating single stranded DNA (ssDNA) fragments, until it reaches a specific recognition motive called Chi-site (Dillingham and Kowalczykowski, 2008). Since Chi sites are a lot more common on the chromosome compared to invasive elements, processing by RecBCD would therefore generate less degradation products that could serve as substrate for spacer acquisition from the chromosome (Levy et al., 2015). Additionally, because double stranded breaks often are the result of stalled replication forks, and due to their high copy number replication forks are more abundant on plasmids, this would lead to a larger number of plasmid derived substrates (Levy et al., 2015). Furthermore, phage genomes often exhibit unprotected dsDNA ends upon entering the cell, as well as before packaging, further biasing RecBCD-mediated generation of spacer substrates towards foreign DNA (Levy et al., 2015). In support of this the fraction of spacers originating from the chromosome increased when RecBCD components were deleted. Moreover, spacer acquisition was more pronounced in areas between double stranded breaks and Chi sites (Levy et al., 2015). As the Cas1-2 complex has been shown to prefer dsDNA substrates (Arslan et al., 2014; Nuñez et al., 2015; Wang et al., 2015; Nuñez et al., 2015), it has been proposed the ssDNA fragments generated by RecBCD reanneal to form dsDNA substrates for spacer acquisition (Amitai and Sorek, 2016; Jackson et al., 2017).

**Primed adaptation**

Primed adaptation makes use of pre-existing spacers to acquire additional spacers targeting the same, or similar, invaders (Swarts et al., 2012; Datsenko et al., 2012). This is thought to enhance the immune response and prevent escape mutants from emerging. In the type I-E system the new spacers acquired by priming typically target the same strand as the original spacer (Swarts et al., 2012). Priming requires all components of the CRISPR-Cas system (Datsenko et al., 2012). The interference machinery is likely responsible for the generation of prespacer donor fragments, while Cas1 and Cas2 are still required for actual integration to take place. Conformational changes in target bound Cascade seems to result in either direct recruitment of Cas3 and typical interference, or Cas1-Cas2 mediated recruitment of Cas3 and primed adaptation (Blosser et al., 2015; Redding et al., 2015; Hayes et al., 2016).
Cas3 generates 30-100 nt single stranded degradation products in vitro, these fragments are enriched for PAM sequences at their 3’ end and likely re-anneal to form partially double stranded substrates for Cas1-Cas2 (Künne et al., 2016). Both escape mutants as well as targets evoking a productive interference response can trigger primed adaptation (Swarts et al., 2012; Datsenko et al., 2012; Semenova et al., 2016; Staals et al., 2016). It appears spacers that do not raise an effective interference response trigger more primed spacer acquisition than interference proficient spacers. This is probably due to the prolonged presence of escape mutants in the cell allowing for more prespacer substrate production and eventually primed spacers integrated (Semenova et al., 2016; Künne et al., 2016).

Expression
The expression stage involves transcription and translation of the cas genes, as well as transcription of the CRISPR-array and processing of the resulting precursor crRNA transcript (pre-crRNA) into mature crRNAs. These crRNAs are combined with the effector Cas proteins to form the active ribonucleoprotein complex.

Regulation
There can be costs associated with expression of CRISPR-Cas systems (Westra et al., 2015). For example, there is the risk of lethal self-targeting if chromosomally derived spacers are acquired (Stern et al., 2010). But uptake of mobile genetic elements can also increase fitness, for example if they carry antibiotic resistance or virulence factors (Jiang, et al., 2013). Due to selection, some microbes might therefore loose or inactivate their CRISPR-Cas system, or alternatively regulate expression of CRISPR-Cas components. Regulation of CRISPR-Cas expression has been shown to occur across a variety of hosts and systems (Patterson et al., 2017).

Expression of the E. coli K12 cas genes is controlled by the global transcriptional regulators H-NS and LeuO (Pul et al., 2010; Pougach et al., 2010; Westra et al., 2010). Under laboratory growth conditions native cas gene expression is not enough to provide immunity to phage infection (Westra et al., 2010). H-NS silences transcription of AT-rich, often horizontally acquired, DNA. As CRISPR-Cas systems are thought to have originated from mobile genetic elements this might be a reason for its silencing by H-NS (Pul et al., 2010). It has been proposed that AT-rich invaders might titrate H-NS away from the CRISPR-Cas promotors, coupling infection to de-repression (Pul et al., 2010). However, H-NS is abundant and it remains to be demonstrated that invading elements would be adequate to displace enough H-NS away from the CRISPR-Cas locus (Patterson et al., 2017). What has been shown is that H-NS repression can be relieved by the transcriptional regulator LeuO (Westra et al., 2010). There are indications of a potential link between membrane stress
as a result of phage infection and CRISPR-Cas induction (Ratner et al., 2015; Patterson et al., 2017). In E. coli, membrane stress due to the disrupted folding and translocation of newly synthesized proteins has been shown to induce CRISPR-Cas expression, possibly via the BaeSR two component system (Perez-Rodriguez et al., 2011). Furthermore, stability of Cas3 at 37°C seems dependent on the heat shock chaperone HtpG (Yosef et al., 2011; Majsec et al., 2016). HtpG has been shown to be upregulated in E. coli during phage PRD1 infection (Poranen et al., 2006).

**Processing**

Transcription of the E. coli CRISPR-array is driven from a promoter inside the upstream leader sequence and generates a long precursor transcript known as pre-crRNA (Brouns et al., 2008; Pul et al., 2010; Pougach et al., 2010). In the type I-E system Cas6e is responsible for the processing of the pre-crRNA into mature crRNAs by endonucleolytic cleavage within the repeat sequence (Brouns et al., 2008). This results in 61 nt long crRNAs that have a 5’ handle consisting of 8nt derived from the repeat, followed by the entire spacer sequence, and the remainder of the repeat forms a hairpin at the 3’ end (Brouns et al., 2008; Jore et al., 2011). Cas6e specificity is determined by recognition of both repeat structure and sequence (Gesner et al., 2011; Sashital et al., 2011). After cleavage Cas6e remains associated to the crRNA (Sashital et al., 2011).

Type I and type III systems all employ an endoribonuclease, either stand alone or in a complex, for processing of the pre-crRNA, similarly as has been described for the type I-E system above (Charpentier et al., 2015). In contrast, type II systems employ a trans-acting small RNA (tracrRNA) which base-pairs with each repeat, leading to cleavage of the double stranded RNA by RNase III (Deltcheva et al., 2011).

**Interference**

In the interference stage the CRISPR-Cas effector complex consisting of crRNAs associated with Cas proteins, surveys the cell for invaders. Target identification occurs by base-pairing between the crRNA and the complementary sequence in the target nucleic acid, called protospacer, as well as PAM recognition (see below). This triggers degradation of the target by specific Cas nucleases.

**Protospacer adjacent motif**

As the crRNA spacer is not only complementary to the target protospacer, but per definition also to its template in the CRISPR-array, differentiating self from non-self is critical for CRISPR-Cas systems. In type I CRISPR-Cas systems this is achieved by detection of a conserved sequence at the 3’ end of the
protospacer, called the protospacer adjacent motif or PAM (Mojica et al., 2009; Westra et al., 2013). Protein-mediated recognition of the PAM activating the CRISPR-Cas machinery for interference (see below ‘surveillance and interference’) (Sashital et al., 2012; Westra et al., 2013; Hayes et al., 2016). Key feature of the PAM is thus that its sequence differs from the corresponding sequence in the repeat. In silico analysis of sequences flanking the protospacer revealed a CWT (W = A or T) consensus PAM in E. coli (Mojica et al., 2009). Note that there is no consensus in the field regarding how the PAM sequence is indicated, nevertheless throughout this thesis the PAM is always given in the 5’ to 3’ direction as it would occur on the target strand. In vivo studies in E. coli confirmed the importance of the PAM for phage resistance and showed Cascade accepted CAT, CTT, CCT, and CTC PAMs (Semenova et al., 2011; Westra et al., 2012; Westra et al., 2013). A recent crystal structure of E. coli Cascade bound to foreign DNA provided structural insights into PAM recognition and suggested a more general CHH (H = A, C or T) PAM (Hayes et al., 2016). The least preferred PAM according to this structure would be CGG, which is incidentally the ‘PAM’ of the spacer in the CRISPR-array (Hayes et al., 2016).

To guaranty that functional spacers are inserted in the host CRISPR locus the PAM is also recognized by the Cas1-Cas2 complex during adaptation (see ‘Adaptation’). However, the exact motifs recognized by Cas1-Cas2 during adaptation and Cascade during interference, while likely overlapping, are not necessarily identical (Shah et al., 2013).

**Cascade complex formation**

The effector complex of the type I-E CRISPR-Cas systems is called Cascade. Cascade consists of a core complex including the 61 nt crRNA, Cas5e, Cas6e, and six Cas7 subunits, as well as a Cse1 and two Cse2 subunits which are more loosely associated with the complex (Brouns et al., 2008; Jore et al., 2011; Wiedenheft et al., 2011; Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014). The shape of the Cascade structure has been likened to that of a seahorse (Jore et al., 2011). According to this analogy, the Cas6e subunit bound to the 3’ stemloop of the crRNA is at the head of the complex. Six Cas7 subunits bind along the crRNA forming the backbone of Cascade. This divides the spacer sequence of the crRNA in 6 segments, each with 5 nucleotides available for interaction and the sixth base flipped out and bound by a Cas7 subunit (Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014). Consequently, mismatches between the spacer and protospacer are tolerated at each sixth position. The 5’ handle of the crRNA at the bottom of the complex is recognized by Cas5e this may function to limit Cas7 oligomerization, as well as provide a platform for Cse1 interaction (Jackson et al., 2014). The crRNA is thus fully protected by Cas protein binding. The two Cse2 subunits represent the belly, and do not contact the crRNA.
**Surveillance and interference**

Scanning of DNA occurs through non-specific interactions between Cascade and the DNA, and are mediated by the Cse1 subunit. Cse1 is also responsible for PAM recognition (Sashital *et al.*, 2012; Hayes *et al.*, 2016). Recognition of the PAM by Cse1 destabilizes the dsDNA and enables crRNA-protospacer hybridization (Hayes *et al.*, 2016). Upon successful seed pairing strand invasion proceeds through the entire spacer sequence, leading to R-loop formation. The non-target strand that is displaced during R-loop formation is stabilized by the two Cse2 subunits (Hayes *et al.*, 2016). R-loop formation induces conformational changes within Cascade, exposing an interaction site on Cse1 that allows for recruitment of Cas3 (Westra *et al.*, 2012; Hochstrasser *et al.*, 2014; Mulepati *et al.*, 2014). Cas3 exhibits both ssDNA nuclease as well as ATP-dependent helicase activity (Sinkunas *et al.*, 2011). The nuclease domain nicks the non-target strand, after which helicase activity is activated and the combined action of nuclease and helicase domains continues to degraded the target DNA in the 3’ to 5’ direction (Sinkunas *et al.*, 2013; Mulepati and Bailey, 2013).

**CRISPR-Cas applications**

Apart from its biological functions in prokaryotes, CRISPR-Cas systems have gotten a lot of attention as biotechnological tools (reviewed in (Jiang and Marraffini, 2015; Sternberg and Doudna, 2015)). Most of these tools utilize the programmable sequence specific DNA recognition of the interference complexes. The best known is CRISPR-Cas9 mediated genome editing in eukaryotes. Cas9 is especially suitable for genome editing as it contains both target recognition and nuclease activity in a single protein. Fusing the crRNA with the tracrRNA into a chimeric single guide RNA (sgRNA) reduces the components required for target recognition and interference to just two, (Jinek *et al.*, 2012). Importantly, Cas9 creates a single double stranded break upon target recognition (Jinek *et al.*, 2012), in contrast to Cas3 which progressively degrades DNA as discussed previously (see ‘interference’). Heterologous expression of Cas9 and a sgRNA designed have complementarity to the gene of interest leads to a Cas9 induced double stranded break at the target locus (Jinek *et al.*, 2013; Cong *et al.*, 2013; Mali *et al.*, 2013; Cho *et al.*, 2013). The double stranded break is subsequently repaired by the endogenous DNA repair pathways of the eukaryotic host, resulting in genetic modifications (Jinek *et al.*, 2013; Cong *et al.*, 2013; Mali *et al.*, 2013; Cho *et al.*, 2013). If the lesion is repaired via non-homologous end joining this may result in gene disruption, because the process is error prone, causing gene crippling frameshifts by the insertion or deletion of several bases. If a donor DNA molecule is provided sharing homology with the targeted sequence, or if another allele is present in the cell, homology directed repair can be used to repair the break. Although
usually less efficient than non-homologous end joining this enables specific changes such as introduction of epitope tags or a mutation of interest. CRISPR-Cas9 has now been used for genome editing in virtually every model organism and cell-line (Jiang and Marraffini, 2015; Sternberg et al., 2016). There are some limitations, such as the need for a PAM sequence within the target region of interest, and the efficiency of editing varies widely depending on cell type and delivery method (Sternberg and Doudna, 2015). Also, as with any gene editing technique, off target effects are a worry, and efforts are made to reduce these (Sternberg and Doudna, 2015). Recently also Cas12a (former Cpf1), another CRISPR class 2 nuclease, has been employed for gene editing in Eukaryotes (Zetsche et al., 2015; Zetsche et al., 2017). Cas12a does not make use of a tracrRNA or host-factors for processing, therefore the crRNAs can be provided as a long pre-crRNA allowing for easy multiplexing (Zetsche et al., 2015; Zetsche et al., 2017). The unprecedented potential CRISPR-Cas9 offers for genome editing in human and non-human genomes has also prompted an ethical debate, especially regarding human germ-line editing (Baltimore et al., 2015; Bosley et al., 2015; Camporesi and Cavaliere, 2016). As well as in agriculture. Since gene editing by Cas9 does not leave any DNA of foreign origin, are CRISPR-edited plants and animals still genetically modified organisms (GMOs) (Camporesi and Cavaliere, 2016)? In Sweden and several other countries, they are not classified as such. The possibility of eradicating disease vectors through a so called gene drive utilizing CRISPR-Cas9 technology holds a lot of promise for improvement of human health, but might have unpredicted consequences concerning the ecosystem (Camporesi and Cavaliere, 2016).

As bacteria do not have the same capabilities as eukaryotes regarding DNA repair, the primary outcome of CRISPR-induced genomic DNA damage is death. Cas9 applications in bacteria utilize this to selectively remove cells with a specific genotype from the population. This can for example be used in combination with lambda red recombineering, to remove the unchanged cells from the population (Jiang, et al., 2013). Most notably Cas9 has been applied to target antibiotic resistance and virulence genes, effectively eliminating those bacterial species, and plasmids, that carried those markers from the population (Gomaa et al., 2014; Bikard et al., 2014; Citorik et al., 2014).

Other applications make use of the programmable sequence specific targeting capabilities of Cas9 but not its nuclease activity, employing the catalytically inactive dCas9. dCas9 has been used in bacteria to downregulate expression by targeting it to promoter regions where is sterically hinders RNA polymerase from initiating transcription (Qi et al., 2013; Bikard et al., 2013). The same approach has been applied to the larger Cascade complex as well ((Luo et al., 2015) and paper V). The titratability of dCas9 mediated gene silencing has been exploited for functional analysis of essential genes by creating partial knockdowns that did not affect viability (Peters et al., 2016). By fusing dCas9 to transcriptional repressors and activators gene expression can be controlled
in eukaryotes as well (Gilbert et al., 2013; Maeder et al., 2013; Perez-Pinera et al., 2013). Another interesting application is imaging of DNA loci in live cells by a dCas9-GFP fusion (Chen et al., 2013).
Present investigation

The work included in this thesis can roughly be divided into the parts concerning sRNA-mediated regulation, and those related to the CRISPR-Cas adaptive immune system. As pointed out at the beginning, the use of RNA molecules to find target nucleic acids within the cell is a common denominator. As outlined in the introduction, sRNAs are important omnipresent post-transcriptional regulators in bacteria. They regulate through direct base-pairing interaction with their target mRNAs, and the trans-encoded variety in enterobacteria – almost without exception – requires the RNA chaperone Hfq for function. Paper I revolves around the different roles that the RNA chaperone Hfq can play as an RNA matchmaker. It is important to consider not only the mechanisms, but also the physiological impact of sRNA-dependent regulation of gene expression. Paper II explores the specific question of how bacteria exploit regulatory RNAs, or RNA elements, to create the subpopulation of transiently antibiotic-tolerant cells termed persisters.

Papers III and IV, related to the prokaryotic CRISPR-Cas adaptive immune system, could be described as method development. The first stage of CRISPR-Cas immunity is adaptation, the creation of molecular memory by integrating spacer sequences originating from foreign genetic elements into the CRISPR-array on the host chromosome. To facilitate studies on CRISPR-Cas adaptation, paper III describes a highly sensitive method to detect the occurrence of spacer integration events in single cells. The specificity of the immune system comes from base-pairing of processed RNAs derived from the CRISPR-array, the crRNAs, to target DNA or RNA, resulting in cleavage by the effector CAS protein(s). In paper IV we utilize the specificity of the *E. coli* CRISPR/Cas system to inhibit transcription of a gene of interest by targeting a catalytically inactive effector complex to a promoter region of choice.

**Paper I**

In this paper, we identified the DGC YdaM, which functions within the biofilm regulatory cascade, as a target of the sRNAs OmrA and OmrB. Both sRNAs inhibit *ydaM* expression by direct competition with initiation ribosomes, and require Hfq to do so. Interestingly, our results suggest a non-canonical mechanism for Hfq-mediated *ydaM*-OmrA/B base-pair formation. Instead of serving as an RNA interaction platform, Hfq restructures *ydaM* to
enable sRNA binding. Simultaneous interaction with both RNA binding partners is not required.

**ydaM as an OmrA and OmrB target**

Our lab has previously identified the master regulator of biofilm formation *csgD* as an OmrA/B target. After mutating the OmrA/B binding sites in both *csgD* and the TF *ompR*, overexpression of OmrA/B still gave white colonies on Congo red plates (indicative of curli negative strains, curli being the key component of the *E. coli* biofilm matrix). Hence, we searched for additional OmrA/B targets in the biofilm regulatory cascade. Computational predictions indicated putative base-pairing between OmrA and OmrB and a site downstream of the *ydaM* start codon. We confirmed that OmrA and OmrB directly inhibit *ydaM* by measuring fluorescence of a *ydaM-gfp* translational fusion in the presence and absence of OmrA or OmrB overexpression. The interaction site was confirmed by mutations. The interaction site is in close enough proximity to the RBS for binding to result in translational inhibition of *ydaM* expression. Indeed, toeprint analysis and *in vitro* translation assays showed translational inhibition by direct competition with initiating ribosomes. By enzymatic probing, we mapped two OmrA/B binding sites on the *ydaM* mRNA. Besides the predicted and subsequently confirmed binding site, there was an additional interaction site further downstream into the *ydaM* ORF. However, as mutation of the first binding site almost entirely abolished regulation, the second site probably does not contribute significantly to regulation.

**Hfq-mediated RNA remodeling key to *ydaM*-OmrA/B duplex formation**

OmrA and OmrB require Hfq for their stability *in vivo* (Holmqvist *et al.*, 2010). It was therefore not surprising that overexpression of OmrA or OmrB does not regulate a *ydaM-gfp* translational fusion in a Δhfq background. However, we observed that in an *in vitro* translation reaction, where RNAs are entirely stable, addition of OmrA/B alone barely had any effect on *ydaM* translation. However, translation of *ydaM* mRNA was nearly abolished when both sRNA and Hfq were present. An explanation for this strong regulatory dependency on Hfq may lie in the *ydaM* mRNA structure. Footprinting data and *in silico* structure predictions (mFold (Zuker, 2003)), suggested that the seed region of the OmrA/B interaction site in the *ydaM* mRNA was partially sequestered in a hairpin structure. Based on the mapped Hfq interaction site we hypothesized that Hfq binding would melt the inhibitory hairpin and thereby facilitate *ydaM*-OmrA/B duplex formation. In line with this hypothesis, introduction of a Hfq-distal face mutant, unable to interact with *ydaM* mRNA, failed to enhance OmrA/B-dependent regulation of *ydaM* in the *in vitro* translation reaction. In contrast, addition of Hfq-prox and -rim mutant proteins, which can still interact with *ydaM* mRNA, resulted in translational inhibition to near wt levels. Both Hfq-prox and Hfq-rim are expected to have reduced capabilities for interaction with most sRNAs, including OmrA and OmrB.
Using mobility shift assays, we confirmed that OmrA does not interact with Hfq-rim, while we detect binding to Hfq-prox. I.e., the Hfq-rim mutant, still capable of promoting regulation, is unable to interact with both RNAs simultaneously and to function as a platform. Thus, this implies an mRNA remodeling-only function for Hfq in ydaM-OmrA/B duplex formation. In order to confirm this, we annealed a DNA-oligo complementary to the Hfq binding site to the ydaM mRNA, thereby forcing the inhibitory structure to open in a Hfq-independent manner. Indeed, the ydaM mRNA with an annealed Hfq-mimic oligo was markedly inhibited by the sRNAs in absence of Hfq.

Hfq is generally thought of as a ‘platform’ for RNA interaction, both restructuring and bringing RNAs in close proximity to promote base-pairing. While for several mRNA-sRNA pairs, this is supported (e.g. (Peng, et al., 2014; Dimastrogiovanni et al., 2014)), the case of ydaM and OmrA/B shows that Hfq can also be primarily required for its capability to unfold RNA structures without a need to enhance RNA-RNA interaction by increasing the local concentration. Hfq-mediated RNA restructuring and molecular crowding should therefore be considered as discrete functions, which may be distinct from each other, or act in concert, to enhance duplex formation. If one or both functions are required will depend on the specific characteristics of the sRNA-mRNA pair.

OmrA/B and the control of biofilm formation

By inhibiting translation of two positive determinants for biofilm formation, CsgD and YdaM, OmrA/B contribute considerably to the stabilization of a ‘biofilm-OFF’ state. What is the underlying physiological reason for this? We have data (not included in this thesis) that shows that, in the absence of OmrA/B, the onset of curli production is earlier in the growth-curve. This raises the possibility that the delay in biofilm production caused by OmrA/B is important for keeping the opposing pathways of flagella and biofilm formation inversely regulated. Another clue in this direction is that OmrA/B inhibit translation of the anti-Sigma factor FlgM (Hoekzema et al. in preparation). In doing so, OmrA/B positively affect motility, since FlgM is a negative determinant for flagella production. Based on these interactions, OmrA/B would stabilize a flagella-ON/biofilm-OFF program. However, this picture is complicated by the fact that OmrA/B also inhibits expression of flhD. The TF FlhDC is responsible for the activation of class 2 flagella genes (De Lay and Gottesman, 2012). Modulation of FlhDC, YdaM, and CsgD translation under osmotic stress, or other conditions, when omrA and omrB expression is activated via the OmpR/EnvZ two-component system, could be important because assembly of large surface structures such as flagella and curli-fibers might have detrimental effects on fitness or cell survival (Mika and Hengge, 2014). The apparent discrepancy between OmrA/B regulation of flhD and flgM, inhibiting and activating flagellar genes, respectively, might be related
to the different regulons that FlhDC and FliA (the Sigma factor inhibited by FlgM) control; they do not only include flagella genes (Zhao et al., 2007; Fitzgerald et al., 2014). This conundrum certainly merits further investigation.

The expression of csgD is bistable (Grantcharova et al., 2010). Bistability likely arises at the level of csgD transcription, through a YciR-based double negative feedback loop (Lindenberg et al., 2013). Creating phenotypic heterogeneity in a population is thought of as a bet-hedging strategy (see below) (Veening et al., 2008). We have observed that OmrA/B affects the distribution of biofilm ON/OFF cells. OmrA/B thus affect the "amplitude" of bet-hedging in the bistable formation of biofilm. This might be a more common theme for sRNAs involved in bistable networks, as also addressed below for the sRNA antitoxin IstR-1 and TisB-dependent formation of persister cells.

**Future perspectives**

It would be especially interesting to find out whether the primary role of Hfq as RNA remodeler – rather than interaction platform – for ydaM-OmrA/B duplex formation that we observed in vitro also holds water in vivo. For this purpose, the Hfq distal, proximal and rim mutants could be introduced onto the E. coli chromosome, and their effectiveness in OmrA/B-dependent inhibition of ydaM expression investigated by using a ydaM translational reporter fusion. A similar strategy has been used before (Zhang et al., 2013; Schu et al., 2015). However, since the intracellular stability of OmrA/B depends on Hfq (Holmqvist et al., 2010), this may limit conclusions that can be drawn from this type of experiment. Though, as OmrA/B can interact with the Hfq-dist ((Schu et al., 2015), paper I), we can expect them to be stable in a corresponding mutant background. Our model predicts that because Hfq-dist cannot open the ydaM inhibitory structure, and hence ydaM expression should not change in the presence or absence of OmrA/B. As ydaM mRNA levels could be affected by lack of Hfq interaction, differences in translation rates between the wt and Hfq-dist strain must be controlled for. A possible way to circumvent the problem of unstable OmrA/B is by creation of a hybrid RNA, in which the target recognition region of a Hfq-independent sRNA is replaced by the region in OmrA/B important for ydaM recognition. Another complicating factor is that we do not know how the Hfq mutant proteins affect regulation of ydaM by other sRNAs. E.g., RprA also regulates ydaM (Mika et al., 2012). These effects can be minimized by using overexpression of OmrA/B and ydaM transcribed from a heterologous promoter, combined with a mutation in the RprA binding site in ydaM. A different, complementary approach to the question of in vivo relevance would be to open the ydaM inhibitory structure by well-chosen point mutations. This should, if the model is correct, reduce if not eliminate the regulatory requirement for Hfq. This approach is also complicated by low OmrA/B stability in the Hfq deletion strain.

We found that OmrA/B recognize the ydaM mRNA at two distinct but adjacent positions, even though only binding site 1 appears to be required for
regulation. The presence of a second, in terms of regulation non-functional, binding site on the ydaM messenger is somewhat puzzling. It could of course be purely co- incidental, but might have a role that escaped our investigations. Evolutionary conservation can inform on functionality, but ydaM is only present in E. coli and very closely related species. Hence, sequence comparisons do not provide answers. Two-site interaction with one sRNA has been reported before for the lpxR and nhaB mRNAs, but these binding sites were spaced further apart and interaction with either elicited a regulatory response. A possibility is that the adjacent second interacton site functions as a cooperativity-generating binding site. Our data contain some indications that this might be the case. Mobility shift assays suggested that OmrB binds less readily to ydaM mRNA than does OmrA. This correlates with our finding that OmrA is more efficient at binding site 2 compared to OmrB. Moreover, when we added a blocking oligo to site 2, OmrB binding to ydaM site 1 improved considerably. These are indications that binding at site 2 may improve binding at site 1. More experiments are required to determine the functionality of site 2.

Paper II

This paper addresses the question how bacteria utilize regulatory RNAs to create phenotypic heterogeneity, by studying the role of the tisB/istR-1 type 1 TA system in SOS-induced persister cell formation in E. coli.

RNA-based regulation and phenotypic heterogeneity

Phenotypic heterogeneity is the presence of phenotypically different individuals within the genetically identical population. These phenotypically different individuals might have a reduced fitness in the current circumstances, but could be better adapted to the future environment. This survival strategy is called ‘bet-hedging’, and increases fitness of an isogenic population across fluctuating environments (Veening et al., 2008; Grimbergen et al., 2015). Various mechanisms are involved in creating phenotypic heterogeneity, but a key determinant is thought to be stochastic variation, or noise. Noise arises because of random fluctuations in the biochemical reactions in the cell (Kaern et al., 2005; Fraser and Kaern, 2009). Amplification of noise by net positive feedback can result in a bimodal distribution of gene expression, causing some cells to switch into one physiological state while others do not. This is called bistability, where two stable states, or distinguishable phenotypes, are present at the same time within an isogenic population. As discussed before, post-transcriptional regulation by sRNAs shows a threshold linear mode of action that can filter out transcriptional noise, but also experiences ultra-sensitivity near the threshold which might generate heterogeneity (Levine et al., 2007; Levine and Hwa, 2008). This was shown for the B. subtilis sRNA RnaC,
which increases heterogeneity in growth rates by post-transcriptional regulation of the TF AbrB (Mars et al., 2015). Formation of transiently antibiotic-tolerant bacterial persisters is a well-established bet-hedging phenomenon. The type 1 TA-system tisB/istR-I has been implicated in persister cell formation (Dörr et al., 2010). In paper II we investigated how the type 1 TA system tisB/istR-I contributes to formation of a subpopulation of persister cells in E. coli, and in particular the role of the RNA antitoxin IstR-1.

The small hydrophobic TisB toxin localizes to the inner membrane forming pores that could cause loss of membrane potential (Unoson and Wagner, 2008; Gurnev et al., 2012). Using flow cytometry and the membrane potential-sensitive dye DiBAC$_4$ to stain depolarized cells, we found that approximately 50% of wt cells showed tisB-dependent depolarization after treatment with the DNA-damaging antibiotic ciprofloxacin in exponential phase. This type of genotypic heterogeneity, generated in response to environmental change (here DNA damage) rather than arising stochastically, is called responsive diversification (Kotte et al., 2014). The bistable polarization phenotype was not generated at the level of tisB transcription, as a transcriptional fusion of syfp2 to the tisB promoter showed activity in 95% of the cells. Similar results were obtained with a transcriptional fusion to the SOS-induced sulA promoter. Indeed, the threshold for TisB-dependent depolarization is set post-transcriptionally by IstR-1, as shown convincingly by an increase in the depolarized sub-population in a ΔistR-1 strain, whereas IstR-1 overexpression strongly decreases depolarization. Interestingly, the emergence of depolarized cells also occurred earlier in the ΔistR-1 background. Likewise, the HipB antitoxin of the type II TA system hipA/hipB sets the threshold for HipA-dependent persister formation (Rotem et al., 2010).

So, induction of tisB by DNA damage generates two distinct subpopulations, polarized and depolarized cells, depending on IstR-1 levels. In wt E. coli, the depolarized population consists of about 50% of the cells. However, only 0.01%-0.02% of cells also become TisB-dependent persisters. We therefore hypothesize that two thresholds are involved. The first one is reached when TisB translation is sufficiently high to form pores in the membrane and depolarize the cell. This threshold is set by IstR-1 and the 5’ structural element (see below). To become a persister, a second threshold has to be crossed. Perhaps higher levels of TisB reached in only a subset of cells directly drives dormancy, or alternatively high levels of TisB and membrane damage induce downstream genes that lead to persister cell formation. Another attractive possibility is that only a small subset of cells stochastically produces TisB rapidly enough to induce a persistent state before the detrimental effects of DNA damage become lethal, killing the remaining cells

**Role of the structured 5’ UTR**

To ensure tight regulation of toxin expression, many type 1 TA systems contain additional regulatory elements besides sRNA antitoxins. In the case of
tisB, the 5’ UTR of the primary +1 transcript contains an inhibitory structure that conceals a ribosome standby site required for translation (Darfeuille et al., 2007). Processing into the +42 transcript releases the standby site, which allows for standby binding of 30S subunits, making translation possible despite the structured RBS. To study the contribution of the tisB structured 5’ UTR to regulation and persister cell formation, we deleted the first 41 nt of tisB (Δ1-41), so that transcription of tisB immediately results in the translationally active +42 transcripts.

The Δ1-41 strain showed a considerable subpopulation of depolarized cells already at low concentrations of ciprofloxacin (0.1x the minimal inhibitory concentration) that do not trigger induction of the SOS-response. In the double ΔistR-1 Δ1-41 mutant, depolarization was increased 18-fold compared to wt. Persister levels reflected the results obtained for depolarization with an 11- and >100-fold increase compared to wt for the Δ1-41 and ΔistR-1 Δ1-41 deletion strains, respectively. This leads us to conclude that the 5’ inhibitory structure complements IstR-1 to prevent premature tisB expression. Interestingly, persister levels for the ΔistR-1 single deletion were increased only 4-fold compared to wt, indicating that the 5’ UTR inhibitory structure is the more important element for limiting persister cell formation, at least under these conditions.

Pre-induction of the SOS response by incubating the cells with a low concentration of ciprofloxacin prior to challenge with a high ciprofloxacin dose increases persister levels (Dörr et al., 2009). This is in line with TisB-dependent persisters forming in response to environmental signals as is the case with responsive diversification. Interestingly, deletion of both IstR-1 and the 5’ UTR structure causes a strong increase in persister formation regardless of pre-treatment. This indicates a stochastic production of TisB-dependent persisters when both RNA regulatory elements are absent, and may indicate that stress-induced persister formation increases fitness compared to a purely stochastic process.

**Exit from the persistent state**

In order for bacterial persisters to repopulate the environment when conditions are favorable once again, cells need to exit the persister state. Automated imaging of persister assays using the ScanLag method (Levin-Reisman et al., 2010; Levin-Reisman et al., 2014) showed that the ΔistR-1 Δ1-41 double mutant strain not only produces more persisters, but also exhibited a prolonged persistence time, indicated by colonies appearing later compared to the wt strain. When exponential phase cells were plated without prior antibiotic treatment, both wt and single deletion strains showed a narrow distribution of colony appearance time, whereas the double mutant produced a considerable number of late appearing colonies. A possible explanation for this could be increased TisB toxin levels, which is supported by induction of +42 tisB ex-
pression from the arabinose-inducible pBAD promoter resulting in late appearing colonies. This is similar to what was seen for the HipA toxin (Rotem et al., 2010).

Future perspectives
A major remaining question concerns the second threshold that we propose as a step before persister emergence. Single cell analyses with reporter gene fusions to tisB would be extremely helpful to further study TisB-dependent persister cell formation. However, our attempts to create such reporters so far have been unsuccessful, since all fusions lost toxicity and/ or regulation (B. A. Berghoff unpublished). Also, the search for downstream genes that could report on tisB, has not yet resulted in very promising candidates ((Berghoff et al., 2017) and B. A. Berghoff unpublished). TisB N-terminal His or FLAG tag fusions do not lose toxicity (Vogel et al., 2004; Unoson and Wagner, 2008). Therefore, a possible strategy could be to use fluorophore conjugated antibodies recognizing these epitopes. Normally, one would need to fix and permeabilize the cells in order for the antibody to recognize the antigen. However, it has been shown, for the similar pore forming protein TatA, that the N-terminus may be exposed to the outside (Walther et al., 2013). It is thus conceivable that TisB can be detected using a simpler surface-staining protocol. This approach might therefore allow for single cell analysis of TisB expression while retaining toxicity. For example, TisB expression could be correlated with growth retardation using a plasmid-borne, inducible fluorescent reporter, as has been done previously (Helaine et al., 2010). Sorting cells expressing high levels of TisB using flow cytometry, and subsequent transcriptomic and proteomic analysis of those cells, could reveal whether TisB induces downstream genes that confidently can be interpreted as causal for persister cell formation.

Paper III
During the adaptation stage of CRISPR-Cas immunity new spacers, derived from mobile genetic elements, are integrated in the CRISPR-array. This creates a ‘memory’ of the invader which is used in the interference stage to specifically recognize and degrade invaders. In this paper, we describe the development and characterization of a fluorescent reporter for the sensitive detection of spacer integration in single cells. We subsequently use this reporter to quantify integration events during growth, as well as investigate which sequence elements are responsible for the preferred integration in the CRISPR-II array of E. coli.

Methods to study adaptation in vivo
Several methods to study spacer integration in vivo had already been described before we developed our reporter. The most commonly used method is PCR
amplification of the CRISPR-array with one primer annealing in the leader and the other within a downstream spacer. An integration event will expand the CRISPR-array at the leader proximal end and therefore result in a larger PCR product, which can be detected by gel electrophoresis (Yosef et al., 2012). When this assay is performed on a mixed population, PCR products representing both the parental (unexpanded) array and the expanded arrays will be detected. While easily applied, PCR has drawbacks, such as amplification bias, and variable results depending on how much template was provided in the PCR reaction, as well as how much of this reaction was used for analysis. The sensitivity of this method is low, with a detection limit of 0.4% expanded arrays (Yosef et al., 2012). Higher sensitivity (detection limit of 0.01%) can be achieved when PCR products, or total DNA extracts, are analyzed by high throughput DNA sequencing methods (Sternberg et al., 2016). This has the added advantage of also providing the spacer identities (Savitskaya et al., 2013; Yosef et al., 2013; Levy et al., 2015). However, these methods are costly and time consuming, and if a PCR product is used, still suffer from the same PCR biases.

Other assays look for phage survival (Datsenko et al., 2012; Ivančić-Baće et al., 2015), or plasmid curing (Swarts et al., 2012), after prolonged induction of the entire CRISPR-Cas system. To confirm spacer insertion the CRISPR-arrays of selected clones are PCR amplified and can be sequenced if desired. However, since such methods are based on functionality of the spacer, these assays miss integration of non-functional as well as self-targeting spacers, as these would in most cases be lethal. An elegant reporter avoids this problem by utilizing an out-of-frame antibiotic resistance marker, which is moved into frame by expansion of the array with one spacer-repeat unit of 61 bp (Díez-Villaseñor et al., 2013). This reporter can thereby positively select for clones that have successfully integrated a spacer, without the need for spacer functionality or interference machinery (Díez-Villaseñor et al., 2013).

With exception of high throughput DNA sequencing these methods are not suitable for quantification of spacer integration. PCR amplifying the CRISPR-array of a bacterial population after a spacer acquisition assay can tell whether spacer integration occurred (in > 0.4% of the cells) or not. Comparison of the signal intensities could provide an idea if more or less integration occurred. However, as spacer integration is a rare event, the resolution provided by imaging of PCR products on a gel is not sufficient to reliably quantify differences. And while positive selection by phage survival, plasmid curing, or the reporter from Díez-villaseñor et al. can detect rare integration events, they too do not readily give reliable integration frequencies. Or at least not in a high throughput manner, as this would require establishing the total number of viable cells in the culture, as well as confirming expansion of the array by colony PCR of a statistically relevant number of colonies. We therefore set out to develop an easy to use method that could rapidly and reliably quantify spacer integration events in a population.
A fluorescent reporter for spacer integration

We based our reporter on the reporter from Díez-villaseñor et al. described above, but instead of an antibiotic resistance marker we used yellow fluorescent protein (Yfp) to report on integration. A fluorescent reporter allows for rapid detection of spacer integration by various fluorescent detection methods. Most importantly, the fluorescent reporter combined with flow cytometry allows for single cell detection of spacer integration events, meaning we can reliably quantify spacer integration events. We also decided to place our reporter array on the chromosome rather than on a plasmid, thereby avoiding problems arising from integration occurring on one copy of the plasmid but not on the others.

Our reporter looks as follows: A partial CRISPR-array was placed in-between the start codon and ORF of yfp in such a fashion that the yfp ORF is out of frame and expansion of the CRISPR-array by one spacer repeat unit (61 bp) moves it into frame. Hence, no Yfp is produced unless a successful integration event took place. It should be noted that not all integration events will lead to fluorescence. This is because in some cases spacers of a different length might be integrated, which would not move yfp into frame. Moreover, insertion of spacers containing in frame stop codons will result in transcription termination before Yfp is translated. Considering these factors Díez-villaseñor et al. calculated the probability of detecting spacer integration to be 0.59 (Díez-Villaseñor et al., 2013). The partial CRISPR-array consists of part of the leader including all sequence elements previously determined to be necessary for integration (Yosef et al., 2012; Nuñez et al., 2016), and one spacer flanked by two repeats. Expression is driven by a constitutive promoter. We constructed two versions of the reporter, one with the leader and spacer sequence of the E. coli CRISPR-I array, and another with those of the CRISPR-II array. Reporters were integrated into the galK locus of an E. coli MG1655 strain deleted for all cas genes and both CRISPR-arrays, carrying a Cas1 and Cas2 over-expression vector. Therefore, our assays report on naïve integration only. As negative control the same strain carrying a plasmid with catalytically inactive Cas1 was used.

Reporter characterization

To test whether spacers could be integrated in the reporter array we performed an acquisition assay followed by detection of spacer integration by the established PCR method. As expected integration was observed for both the CRISPR-I and the CRISPR-II reporter. By plating a culture after an acquisition assay fluorescent colonies could be observed. We performed colony PCR on selected fluorescent and non-fluorescent colonies, and found that all fluorescent colonies, but none of the non-fluorescent colonies, had expanded CRISPR-arrays. We can therefore conclude that spacers can be integrated in
the reporter array and that fluorescence is the result of a successful integration event.

To determine the detection limit we constructed a ‘pre-expanded’ positive control by insertion of 61 bp random sequence without stop codons into the first spacer. This positive control was then mixed at set ratios with the negative control (unexpanded array with inactive Cas1), and analyzed by the classical PCR method as well as flow cytometry. Using PCR we could detect 0.5% expanded arrays, which is in good agreement with what the previously reported detection limit of 0.4% expanded arrays (Yosef et al., 2012). However, no distinction could be made between the signal intensities of samples with 0.5, 0.75 and 1% expanded arrays, this method can therefore not reliably quantify small differences. When the samples were analyzed by flow cytometry we could see a clear separation between the Yfp- and Yfp+ populations, gates were set so no more than 0.01% of Yfp+ events were measured in the negative control. This way we were able to reliably detect as little as 0.05% of expanded arrays using flow cytometry. We therefore conclude that detection of spacer integration using our fluorescent reporter in combination with flow cytometry analysis is ten-fold more sensitive than detection by PCR.

Applying the fluorescent reporter

We performed an acquisition assay with the CRISPR-I and CRISPR-II fluorescent reporters and found 0.6 and 3.0% Yfp + cells respectively. It should be noted that, as not all integration events lead to fluorescence (see above), the percentage of fluorescent cells does not provide the absolute integration frequency. However, as the probability of detecting spacer integration is constant, this does not affect the ability to accurately detect differences in integration frequencies between two samples, in this case CRISPR-I and CRISPR-II. A preference for integration into CRISPR II has previously been reported (Datsenko et al., 2012). While the leader region differs substantially between CRISPR-I and CRISPR-II, the repeat only has one point mutation. In order to investigate which sequence elements were responsible for the observed difference in integration frequencies we constructed minimal reporter arrays consisting of the leader and one repeat from either CRISPR-I or CRISPR-II. We constructed all four possible combinations of leader and repeat sequences, and quantified integration into each of them. Both hybrid arrays exhibited an intermediate integration frequency, suggesting that both the leader and repeat contribute to the higher adaptation frequency in CRISPR-II. Further mutational studies are required to determine more precisely which nucleotides are important.

To further demonstrate the applicability of this fluorescent reporter we measured integration frequencies over time during an acquisition assay. A sharp increase in Yfp+ cells was observed in late exponential phase, continued throughout early stationary phase to plateau in late stationary phase. No increase in Yfp + cells was observed in exponential phase. Therefore, under our
experimental conditions adaptation occurs in late-exponential/early-stationary phase. While this growth phase dependence could be biologically relevant, it is also possible that this reflects Cas1 and Cas2 induction kinetics.

**Future perspectives**

This fluorescent CRISPR adaptation reporter could be used to investigate host factor requirements by introducing deletions of candidate genes. Additionally, the general idea of our reporter can be expanded on and adapted for other purposes, and other CRISPR-Cas systems (provided the spacer-repeat unit cannot be divided by three). In our current setup where Cascade and Cas3 are absent only naïve adaptation takes place, while this has the advantage that we do not lose cells due to lethal self-targeting, it would be interesting to look at primed adaptation as well. This would require re-introduction of Cascade and Cas3, or introducing the reporter array into a Δhns strain. The resulting strain would as a result also be proficient in interference, which could allow for the simultaneous detection of integration and interference. A fluorescent reporter for single cell analysis of interference/target clearance, in the form of a target plasmid carrying a constitutively expressed (green fluorescent protein) *gfp* gene, has been successfully used before (Xue *et al.*, 2015; Xue *et al.*, 2016). As emission wavelengths of Yfp and Gfp are very similar it is difficult to separate their signals. It would therefore be better to switch the marker of one reporter for another fluorescent protein to allow for simultaneous and differential detection of both adaptation and interference. A possible problem with detection of primed adaptation is that a second spacer integration event would move the reporter out of frame. As primed adaptation is more efficient and regularly results in multiple integration events, data could be hard to interpret. With our current setup, we can only study adaptation in one array at a time. It is possible that having the option for integration in multiple arrays makes a difference, for example in the case of the CRISPR-I and CRISPR-II arrays. Therefore, it could be of interest to construct a dual array readout using another compatible fluorescent reporter.

**Paper IV**

Precise and efficient molecular tools for the control of gene expression are useful for biotechnological application as well as basic research. In this paper, we demonstrate that the *E. coli* type I-E CRISPR-Cas effector complex Cascade can be used for efficient, long lasting and programmable gene silencing.

**The setup**

We used plasmid based expression of Cascade in an *E. coli* Δcas3 strain. Deletion of *cas3* is required so no degradation occurs upon target DNA recognition by Cascade. A second plasmid with partial CRISPR-array consisting of
the transcribed part of the leader and a spacer flanked by two repeats supplies
the crRNA that guides Cascade to the desired target sequence. Multiplexing
can be achieved by adding more spacers to the CRISPR-array.

**Cascade can successfully silence gene expression**

To test whether Cascade could silence gene expression, we designed four
spacers targeting the promoter of *gfp*, and another four targeting the ORF. Of
each set of four, two spacers targeted the template strand and the two others
the corresponding region of the non-template strand. We then measured Gfp
fluorescence in a microplate reader. We found that targeting of the promoter
region resulted in a 10 to 50-fold decrease in fluorescence. Targeting Cascade
to the ORF was less effective with, on average, a five-fold decrease in fluo-
rescence. Similar results were obtained with a chromosomal as well as plas-
mid borne target. Reduction of fluorescence was stable over time. Flow cy-
tometry showed a similar reduction of fluorescence in all cells, indicating gene
silencing by Cascade is homogenous. To investigate whether Cascade can si-
lence multiple targets simultaneously we introduced a second fluorescent
marker, blue fluorescent protein (Bfp). Both Gfp and Bfp fluorescence could
be simultaneously silenced by expression of crRNAs targeting their promoter.
Silencing was specific, the *gfp* and *bfp* targeting crRNA only reduced Gfp and
Bfp fluorescence respectively.

**Cascade transcriptionally silences target genes**

To confirm that silencing occurs at the level of transcription, *gfp* mRNA levels
were measured by northern blot analysis. This showed a Cascade dependent
absence of *gfp* transcripts in the presence of *gfp* targeting crRNAs but not a
non-targeting crRNA.

**Future perspectives**

A similar study by Luo *et al.* observed higher levels of gene silencing (100-
fold vs 50-fold) (Luo *et al.*, 2015). This discrepancy might be explained by
differences in expression level of the target gene, crRNA or Cascade. Method
of fluorescence detection could also contribute; we performed fluorescence
measurements on a population level in a microplate reader while they per-
formed single-cell measurements by flow-cytometry. Spacer sequence as well
as exact target location can also affect the efficiency of silencing, as was also
observed for the small set of spacers tested in our study. While we observed
more effective silencing when targeting the template strand, Luo *et al.* found
the opposite tendency (Luo *et al.*, 2015). Studies using dCas9 for gene silenc-
ing similarly observed a bias towards the non-template strand (Qi *et al.*, 2013;
Bikard *et al.*, 2013). It is clear that a larger cohort of spacers should be tested
to determine optimal spacer sequence and target region.

Gene silencing using Cascade or dCas9 are comparable in efficiency.
dCas9, as a single catalytically inactive protein, is arguably easier to express
than the multi-subunit Cascade complex. However, the CRISPR-array used by Cascade allows for easy multiplexing, while dCas9 relies on sgRNAs which have to be transcribed or supplied individually. Cascade and dCas9 have different PAM requirements which should be considered when choosing which system to use.

Perhaps the most interesting application for Cascade (or dCas9) mediated gene silencing is the potential to induce knockouts of essential genes without the need for time consuming construction of conditional mutants. Alternatively, expression can be knocked-down to a lower, but non-lethal level, by modulating the degree of silencing. This approach has been used successfully in *B. subtilis* to functionally analyze essential genes (Peters *et al.*, 2016).
Concluding remarks

Since the initial discovery of the regulatory antisense RNAs that control plasmid copy number, and the first chromosomally encoded sRNA MicF, many studies have provided groundbreaking insights into sRNA-based regulation, only a part of which I was able to cover in this thesis. It has become clear that, although rarely essential, sRNAs are important modulators of gene expression that affect all cellular processes. Many, widely varying, regulatory mechanisms are understood in molecular detail. Also, much progress has been made in understanding the RNA chaperoning properties of Hfq, and how they contribute to sRNA-mRNA duplex formation. However, there is still a lot to be learned, as exemplified by the first study included in this thesis. What on the outset seemed to be a straightforward target identification (OmrA/B-ydaM mRNA) and functional characterization (translational inhibition via ribosome exclusion), revealed a new, though perhaps not unexpected, mechanism of action for Hfq. This work showed that Hfq acts primarily to locally restructure the mRNA rather than to serve as a platform for RNA interaction.

Beyond the mechanistics of sRNA-based regulation, it is also important to consider the physiological roles of sRNAs. sRNAs do not act in isolation but are nodes in the complex regulatory networks of the cell, providing its own specific regulatory properties. We studied the E. coli TisB/IstR-1 TA system to explore the role of the sRNA antitoxin IstR-1 and a regulatory structure in the 5’ UTR of the tisB mRNA, in TisB-dependent depolarization and persister cell formation. We found that these regulatory RNA elements are important for balancing the fraction of the population that switches to a persistent state, and are responsible for the responsive diversification of TisB-dependent persister cell formation.

Central to both post-transcriptional regulation by sRNAs and the CRISPR-Cas adaptive immune system is the base-pairing potential of RNA, which allows for specific identification of target nucleic acids. These systems can therefore easily be reprogrammed to target nucleic acids of interest by changing the RNA sequence. We made use of this programmability and repurposed the E. coli type I-E CRISPR-Cas effector complex Cascade for specific transcriptional gene silencing, in a proof-of-principle study.

Before crRNAs can serve as guides for the CRISPR-Cas effector complexes, they have to be integrated into the CRISPR-array. In order to facilitate studies of the intriguing adaptation process, we have developed a fluorescent reporter system for sensitive detection of adaptation events in single cells.

Det är nu vida accepterat att icke-kodande RNA spelar en viktig roll i för post-transkriptionell reglering i alla typer av organismer. I bakterier är små RNA (sRNA) den största gruppen av reglerande RNA. Dessa är viktiga för att modulera genuttryck och är involverade i att reglera många aspekter av bakteriers fysiologi, som till exempel bildande av biofilm och virulens, vilka båda är av klinisk relevans. Jag har undersökt hur mekanismerna bakom post-transkriptionell genreglering sker via sRNA och dessa influerar bildandet av biofilm och persistierande celler, något som är viktigt för bakteriers virulens under t.ex. infektioner.

Biofilm är uppbyggd av populationer av bakterier, där dessa celler samverkar och gemensamt bildar en skyddande hinna (matrix) som består av polysackarider (långa sockerkedjor) och proteinfibrer. Bakterier i dessa biofilmer är mer resistenta mot antibiotikabehandling eller andra stressfaktorer jämfört


I artikel III, beskriver vi en ny metod för att upptäcka förekomsten av sådana integrationer av nya mellansekvenser i enskilda bakterieceller. Vi hoppas.
att den överlägsna känsligheten i vår metod, och möjligheten att upptäcka integration i enskilda celler, kommer förbättra vår förståelse för hur CRISPR-Cas systemet fungerar.


Sammanfattningsvis har mitt arbete lett till en ökad förståelse av hur den post-transkriptionella regleringen av sRNA i bakterier fungerar i stort, och specifikt hur den fungerar i kliniskt relevanta processer då bakterier bildar biofilm och persistenta celler. Jag har dessutom studerat bakteriers immunsystem, CRISPR-Cas, vilket har lett till utvecklingen av två molekylära verktyg. Det första kan med hög känslighet och precision detektera när mellansekvenser sekvenser interageras in i bakteriens kromosom, vilket leder till att skapa ett minne hos systemet. Det andra verktyget har medfört att vi kan tysta ner en specifik gen av intresse genom att blockera transkription.
Alle organismen staan in verbinding met hun omgeving. Dit is gevaarlijk, maar ook noodzakelijk voor het opnemen van de voedingsstoffen die vereist zijn voor alle levensprocessen. Bacteriën hebben daarom een enorm vermogen ontwikkeld om hun omgeving waar te nemen, en snel te reageren op verandering door hun genexpressie aan te passen. Genexpressie is het proces waarbij de genetische instructies, genen bewaard op het chromosoom in de vorm van DNA, worden omgetoverd in functionele producten. De eerste stap in genexpressie is transcriptie, het DNA van een gen wordt gekopieerd naar RNA. Als het gen informatie bevat voor de productie van een eiwit wordt het RNA messenger RNA (mRNA) genoemd. Als dit niet het geval is dan spreken we van niet-coderend RNA. Het mRNA wordt gebruikt door een moleculaire machine genaamd ribosoom als blauwdruk voor het maken van eiwitten. Dit proces van eiwitsynthese wordt translatie genoemd. Genregulatie is het aanpassen van de genexpressie, het aan- en uitzetten van genen onder verschillende omstandigheden. Dit bepaalt de concentratie van een door een gen gecodeerd eiwit of RNA in de cel. Genregulatie kan plaatsvinden op alle niveaus van genexpressie. Regulatie op het niveau van mRNA productie wordt transcriptieregulatie genoemd. Daarbij zijn zogenaamde transcriptie factoren (TFs) betrokken, DNA bindende eiwitten die de transcriptie van mRNA blokkeren of juist activeren. Post-transcriptionele regulatie geschiedt nadat het mRNA al geproduceerd is, maar voordat translatie (eiwitsynthese) heeft plaatsgevonden. Dit proces beïnvloedt de mate waarin het mRNA wordt vertaald naar eiwit en/of de mate waarin het mRNA afgebroken wordt.

Waar vroeger gedacht werd dat vooral eiwitten verantwoordelijk waren voor genregulatie, is het tegenwoordig alom geaccepteerd dat RNA een belangrijke rol speelt in post-transcriptionele gen regulatie in alle organismen. In bacteriën vormen zogenaamde kleine RNAs (sRNAs) de grootste groep van regulators. Ze zijn belangrijk voor het moduleren van genexpressie in alle fysiologische aspecten van bacteriën, zoals bijvoorbeeld de klinisch relevante formatie van biofilm en ook virulentie. In mijn werk heb ik onderzocht hoe sRNAs precies werken en ook hoe ze invloed hebben op biofilm en persister cel formatie. Een biofilm is een multi-cellulaire samenleving, waar veel bacteriën samenkomen en zich omhullen in een extracellulaire matrix (beschermmende laag) van onder andere polysachariden (lange suiker ketens) en eiwitevezels. Bacteriën in biofilms zijn beter bestand tegen antibiotica en andere
stressfactoren, en zijn een klinisch en industrieel probleem. In mijn eerste artikel beschrijven we hoe twee sRNAs, OmrA en OmrB, de translatie van YdaM beletten. YdaM is een eiwit dat de productie van de extravacuolaire matrix componenten activeert. Om dit te doen hebben OmrA en OmrB hulp nodig van een van een mRNA-helper, genaamd Hfq. Dit is niet speciaal, veel sRNAs hebben de hulp van Hfq nodig om genen te reguleren. Maar de manier waarop Hfq OmrA/B helpt bij het reguleren van de ydaM genexpressie is wel nieuw. In de meeste gevallen bindt Hfq zowel het sRNA als het mRNA dat gereguleerd wordt tegelijkertijd, zodat ze makkelijker met elkaar kunnen interacteren. Dit is echter niet nodig voor OmrA/B regulatie van ydaM. In plaats daarvan bindt Hfq het ydaM mRNA en maakt de intermoleculaire structuren in ydaM mRNA strenger losser zodat de OmrA en OmrB sRNAs makkelijker kunnen binden.

In artikel twee laten we zien hoe twee regulerende RNA-elementen bepalen hoeveel cellen in een populatie zogenaamde persister cellen worden. Persisters zijn bacteriën die tijdelijk antibiotica en andere stressfactoren kunnen tolereren omdat ze ‘slapen’. Dit komt omdat de essentiële processen die antibiotica in de war sturen alleen maar essentieel zijn in actieve bacteriële cellen. Als de antibiotic behandeling stopt en deze cellen ‘wakker’ worden, en zich opnieuw gaan reproduceer, kan een nieuwe infectie ontstaan. Binnen een bacteriële populatie worden maar een paar cellen persister. Expressie van toxine-genen draagt bij aan de formatie van deze persister cellen. Toxines beïnvloeden de cellulaire processen en de bacteriën komen in een soort van winterslaap, ze worden persister. Een voorbeeld van zo’n toxine is TisB. De expressie van tisB wordt gereguleerd door een sRNA, Istr-1, en een regulerend RNA-element in het tisB mRNA dat effectieve translatie verhindert. Wij constateerden dat deze twee remmende RNA-elementen het aantal persister cellen dat vormt vermindert. Dit verklaar nog steeds niet helemaal het lage aantal persister cellen, en waarschijnlijk liggen er nog meer elementen en mechanismes ten grondslag aan persister cel formatie.

Organismen moeten ook kunnen reageren op bedreigingen die ze tegenkomen in hun omgeving. Voor ons mensen kunnen dit virussen of bacteriën zijn. Voor bacteriën of archaea bestaat deze bedreiging uit bacteriofagen (virussen die bacteriën infecteren) of andere mobiele genetische elementen. Bacteriën hebben daarom speciale mechanismen ontwikkeld om met dergelijke indringers om te gaan, net zoals wij een immuunsysteem hebben. Een zo’n systeem heet CRISPR-Cas. Het bestaat uit specifieke DNA-sequenties, CRISPR-matrix genaamd (die beschrijven waartegen de bacterie zich moet verdedigen) en Cas-eiwitten (die de daadwerkelijke verdediging tegen indringers uitvoert). CRISPR-Cas is een afkorting van clustered regularly interspaced short palindromic repeats, een beschrijving van hoe de matrix eruitziet. Een CRISPR-matrix bestaat uit korte repeterende DNA-sequenties afgewisseld met unieke tussenliggende sequenties. Net als ons immuunsysteem, kan CRISPR-Cas een
herinnering aan een specifieke indringer creëren en kan zo bescherming bieden tegen toekomstige aanvallen door dezelfde indringer. Het CRISPR-Cas systeem creëert herinneringen aan eerder aangetroffen DNA van bacteriofa gen en andere mobiele genetische elementen, in de vorm van de korte intermediaire sequenties die zijn geïntegreerd in de CRISPR-matrix. In het derde artikel beschrijven we een nieuwe methode om dergelijke integraties in de CRISPR-matrix te detecteren in individuele bacteriële cellen. We hopen dat de superieure gevoeligheid van onze methode, en het vermogen om integratie in afzonderlijke cellen te detecteren, ons begrip van hoe het CRISPR-Cas-systeem werkt, zal verbeteren.

In het effectorstadium van CRISPR-Cas immunitéit, wordt eerst de CRISPR-matrix getranscribeerd in een pre-crRNA, dat vervolgens wordt omgezet in een kortere crRNAs. Deze bestaan uit een tussenliggende sequentie en een repeterende sequentie (afgeschreven van DNA naar RNA). Deze crRNAs vormen samen met één of meer Cas-eiwitten een eiwit-RNA-complex. Dit complex kan met behulp van crRNA vreemde en potentieel gevaarlijke DNA-moleculen te identificeren. Wanneer een crRNA overeenkomt met een sequentie, zullen de Cas-eiwitten het vreemde DNA vernietigen. Door het Cas-eiwit complex te voorzien van een geselecteerd crRNA, kunnen onderzoekers het CRISPR-Cas-systeem sturen naar een specifieke DNA-sequentie. Hiervan wordt bijvoorbeeld gebruik gemaakt in het populaire CRISPR-Cas9 hulpmiddel voor DNA-modificatie. In het vierde artikel gebruikten we bacteriële cellen die een complex van Cas-eiwitten hebben die zijn veranderd zodat ze kunnen binden, maar het DNA niet kunnen degraderen. Als test hebben we deze cellen een crRNA gegeven dat een promotorregio voor een van de bacteriële genen herkent. Omdat het grote Cas-eiwitcomplex andere eiwitten blokkeerde voor binding aan het promotorgebied, werd de transcriptie van dit gen voorkomen. We hebben met dit experiment laten zien dat we het CRISPR-Cas-systeem kunnen gebruiken om de expressie van een specifiek gen van belang te kunnen onderdrukken.

Samengevat heeft mijn werk heeft geleid tot een beter begrip van de achterliggende mechanismen van post-transcriptionele regulatie door bacteriële sRNA in het algemeen, en specifiek hoe sRNAs bijdragen aan klinisch relevante processen zoals biofilm en persister cel formatie. Ik heb ook het bacteriële immuunsysteem CRISPR-Cas bestudeerd, wat heeft geleid tot de ontwikkeling van twee moleculaire hulpmiddelen. De eerste kan met hoge gevoeligheid en precisie detecteren als nieuwe intermediaire sequenties, die het ‘geheugen’ van het CRISPR-Cas systeem vormen, worden geïntegreerd in de CRISPR-matrix. Met het tweede hulpmiddel kunnen onderzoekers specifiek de transcriptie van een interessant gen blokkeren.
I would not be the person I am scientifically, and non-scientifically, if it was not for a whole lot of people. I cannot thank you all, and I cannot thank you enough, but I will make an attempt anyway.

First, my PhD journey would not have started if it was not for my supervisor Gerhart taking a chance on me, thank you for that! These years you have not just cared for the research, but also cared about me as a person, I appreciate that a lot. Thank you for always being there with encouragement and advise. I found your door was always open, and you answered e-mail more or less instantly no matter what. I’m also grateful for the many opportunities to travel and participate in the scientific community, as well as the times you have opened your (summer) house for our group. Magnus L, not many people can say they published two papers with their second supervisor, thank you for those collaborations. I have also enjoyed being a lab assistant on your Microbiology course, I think we worked very well together on that. Thanks also for always including me whenever you had group activities, it has been fun. All other micro PIs, you have all inspired and helped me on this journey to becoming a PhD. The microbiology program is a wonderful supportive and stimulating place to ‘grow up’ as a scientist, and this is largely due to your efforts. Thank you!

The work included in this thesis has not been a solo-effort, I have had the pleasure of collaborating with some great scientists and wonderful people. Lina, thank you for being my friend and helping me get through this long process of becoming a PhD, I couldn’t have done it without you! Sharing a (organized) lab bench with you, the countless scientific and non-scientific discussions, and collaborating on the CRISPR-projects have been a pleasure. And I can’t not mention your superb baked-goods, more cake please! Cedric, thank you for teaching me everything I know about in vitro experiments! Paper number one would not have been without your continuous help and efforts, let’s get it published! I have enjoyed sharing office with you, and our (mostly scientific) conversations. Bork, so many good (and crazy) memories! I still miss having you around in and outside the lab. Thank you for letting me in on your tisB/Istr-1 persister project! Good luck with your own group, and don’t be a stranger! Erik H., thank you for teaching me when I was just starting out as a PhD student. A lot of what I’ve been doing has its origins in your PhD thesis.
Thank you for letting me continue that work, and staying involved with advice and fresh ideas! **Sanna**, thank you for your input, not only on the CRISPR-adaptation reporter, but also all my other projects. I appreciate that you take the time for me, despite your busy schedule! I would also like to thank the master students that have contributed to my projects throughout these years, **Linnea Jonsäll**, **Nienke Jager**, and **Emma Nordenfjäll**.

All you wonderful µ’s, thank you for all the memories, chats, lunches and fika’s that we’ve shared! **Maaike**, I’m so glad you found yourself, hold onto that! **Alisa**, child of two groups. You’ll do great I’m sure of it! **Sofia**, thank you for the girly-chats, heaps of chocolate, encouragement, and science. **Elin**, my dancing-queen. Thank you for your support and friendship throughout these years, I couldn’t have done without it. Ásgeir, it’s been great teaching the micro-course with you, thanks also for all the tech-support. **Jingyi**, I admire your dedication, good luck with your PhD! **Dimitra**, your laugh is infectious, good luck to you too! **Showgy**, thank you for always saying hello to me, even when I still have my grumpy morning face on and most people try to avoid me. **Louise**, good luck with your PhD! The newbies, **Laura**, **Sascha** and **Janna**, good luck with your PhD’s, help each other and don’t forget to have some fun along the way (though I probably do not have to tell you guys that)! **Disa**, another dancing-queen. Thank you for your positivity, and for promptly taking care of things whenever I mentioned a problem to you. **Marcus**, I would have never guessed but you’ve grown on me. Good luck with that Nobel! **Magnus S**, I believe you and **Martin** still have to make good on a promise to bring fika while wearing a white suit with pink tie, or was it a pink suit with a white tie? **Petra**, another one of my Microbiology course students that started a PhD in Micro. Good luck! **Dana**, I’m glad you chose to join micro! Good luck to you too! **Katrin**, thanks for sharing my love of Groningen. Good luck in your new lab! **Anirban**, did you find the reference to your paper? Unfortunately, it was not available in time for me to put the proper one, but it’s in there! **Zhen**, always a surprise, thanks for being a friend! **Jonas**, I wish you all the best! **Bart**, another newby, I’m so glad you came back to micro, good luck! **Jolanda**, thank you for always being happy! **Liis**, welcome to micro! Even though you’re no longer officially part of Micro, our coffee-corner is still regularly honored by your presence, thank you **Mikael**, **Maria-Letizia**, **Jens**, and **Marie**.

My IMBIM friends; **Erik G.**, thank you for teaching me all about cloning and flow cytometry, and letting me use your (self-made) lab equipment. I have enjoyed our talks about life and science over a cup of ‘fin-kaffe’ or icecream over the years. **Erik L.**, Thank you for your help with the MACS and the other Eriks self-made lab equipment! **Marlen**, no longer-IMBIM, welcome to Micro! Thank you for a lot of fun at dissertation parties, especially yours!
“Here’s to nights that turned into mornings, with friends that turned into family.” Petar, through everything you have always been there for me. I’ve felt secure knowing you had my back. Thank you for being such a good friend! Alexandre, thank you for long conversations, making us come to beautiful Provence for a great wedding party, and most of all for the salsa dancing! Paul, I will never forget when you hugged me, instead of the other way around. Should I thank you for the dancing too or will that ruin your image? Ozzy, Umut, Javor, Fabio, Heidor, Ju, Gabi, Luimar, Leonor, Neil, David, Madda, Zaza, Alex T, and all those other beer-club people I cannot remember right now, thank you for the good times!

Ieda and Paulo, obrigado pelo seu apoio e todos os doces enquanto eu estava escrevendo minha tese!

Mijn dank gaat ook uit naar mijn familie, die mij altijd heeft gesteund en liefgehad. Bedankt pap, dat je mij hebt aangemoedigd om te leren, dat je altijd in mij geloofd hebt. Ik had het zonder jouw steun niet gekund! Yoep en Niko, broertjes-lief, ik ben zo trots op jullie! Bedankt dat jullie er zijn. Niko, bedankt voor je hulp bij het maken van de cover, hij is geweldig! José en Jacqueline jullie zijn als een grote zus voor mij, bedankt voor alle steun en de gezellige tijd samen. En de rest van de Hoekzema-clan, jullie zijn allemaal geweldig, ik ben er trost op een Hoek te zijn! De Alonsjes, we zijn met heel veel, maar toch houden we elkaar vast, ik ben heel dankbaar voor die familieband. Er zijn, en er waren, veel mooie, sterke en intelligente vrouwen in mijn level, die mij tot voorbeeld zijn. Tante Lammy, tante Rina, en tante Willemien, jullie liefde en zorg betekend heel veel voor mij. Tante Sarina, jij kon mij begrijpen als geen ander, ik mis je! Oma Hoekzema, Oma Alons lievere oma’s bestaan niet. Ten slotte, mam, jij hebt me in 16 korte jaren genoeg liefde gegeven voor een heel level, en toch mis ik je nog elke dag!

“The best things in life are not planned.” Felipe, you’re the best thing that happened to me during this PhD. Thank you for everything!


Miyakoshi, M., Chao, Y., and Vogel, J. (2015a) Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Current Opinion in Microbiology* **24**: 132–139.


guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic Acids Research* **45**: 367–381.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1637

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)