1

General Introduction

Meishan Liu

Department of Molecular Genetics
University of Groningen
Groningen Biomolecular Sciences and Biotechnology Institute
Groningen, the Netherlands
Lactic acid bacteria

Lactic acid bacteria are a family of bacteria that can convert carbohydrates into lactic acid. This large group of bacteria has been well studied for many decades because of their important roles in food fermentation and human health. The classic early definition of LAB by Orla-Jensen is based on cell morphology, sugar utilization patterns and the range of temperatures at which the bacteria grow. It recognizes LAB as non-sporulating, non-motile Gram-positive cocci or bacilli that are usually anaerobic or facultative aerobic [2] and catabolize sugars mainly into lactic acid [1]. With the development of more advanced molecular tools, the LAB were more elaborately described. For instance, although *Lactococcus lactis* has an aerobic system to assist in its development, it was recently discovered that along with heme source, the presence of oxygen increases *L. lactis* cell growth and the production of proteins and vitamins [3]. 16S rRNA gene and whole-genome sequencing revealed that the LAB belong to the phylum *Firmicutes*, class Bacilli, and order Lactobacillales. The genera can be searched in the NCBI Taxonomy Browser[11]. *L. lactis* belongs to the genus *Lactococcus* in the Streptococcaceae family of the Lactobacillales order; it is the main organism examined in this thesis and will be further introduced hereafter.

Some species of LAB are commensals or pathogens. Pathogenic LAB are found mainly in the Streptococcus family, which can cause life-threatening invasive infections, and Enterococcus genera, which are important causative agents of nosocomial infections [12]. However, a large group of LAB are used in traditional or large-scale industrial food fermentation to the benefit of humans worldwide. The most important genera of food-related LAB are members of Lactobacillus, Lactococcus, Streptococcus, Enterococcus, Pediococcus, Leuconostoc, Oenococcus, Weissella, Carnobacterium and Tetragenococcus [4]. Probiotic LAB, which are intended to provide health benefits when consumed or applied to the body, have been widely used in human health maintenance during the last several decades (Figure 1) [5]–[10].
Figure 1. The benefits of probiotic LAB in human life. 1) LAB affect the intestinal environment of animals, including humans. 2) Some LAB species can inhibit the growth of spoilage bacteria in food through antimicrobial activities or by lowering the environmental pH. 3) In the food industry, LAB play roles in flavor, taste and texture formation. 4) LAB can induce an immune response via binding to specific receptors, and some are used for clinical treatment in various diseases [5]–[10].

**Lactococcus lactis**

The Gram-positive, ovoid-shaped facultative anaerobic bacterium *L. lactis* most likely originates from plant materials and is generally isolated from fermented dairy products. It has been used for millennia to make fermented dairy products. It is one of best characterized LAB. After being reclassified from Streptococcus to Lactococcus, *L. lactis* was divided into three subspecies: *L. lactis* subsp. *lactis*, *L. lactis* subsp. *cremoris*, and *L. lactis* subsp. *hordniae* [14]. It is generally recognized as safe (GRAS), and the subspecies *lactis* and *cremoris* are widely used in the dairy industry as starters for the majority of fermented milk products, such as a large variety of cheeses, quarks, sour cream, butter and buttermilk [15]. One of the main metabolic products of *L. lactis* is lactate, which is an important compound that not only imparts special flavors to dairy products but also contributes to the preservation of fermented products. Currently, benefiting from the development of genetic engineering techniques, *L. lactis* is being explored for broader application in the food industry and in medicine (see below).
As a widely used, industrially important bacterial species, the various uptake mechanisms of different metabolizable sugars in *L. lactis*, the changing of fermentation end products and the energy generated from multiple catabolic pathways have been examined in great detail over the past several decades [30], [31]. Like many other species of LAB, *L. lactis* can convert carbohydrates to lactate by homofermentative or heterofermentative pathways depending on the carbon source and growth conditions. *L. lactis* uses a homofermentative pathway with a favorable carbon source, such as glucose. Lactate is the main product of the metabolic pathway. Alternatively, during aerobic growth or in the presence of less preferred carbon sources, lactate, formate, ethanol, CO₂, acetate, acetoin and other products can be generated as end products of heterolactic fermentation [31]-[34].

*L. lactis* has been employed as a model LAB for the development of genetic engineering tools for this group of organisms. It has a relatively small genome (2.3 Mbp), and the full DNA complement of many strains from various origins has been fully sequenced in the past several decades (source: NCBI). In addition, many genetic, metabolic and protein engineering tools have been introduced and developed for *L. lactis* [35], which has thus become a prime model Gram-positive bacterial alternative to *Bacillus subtilis* and *Lactobacillus plantarum*. 

---

### Table 1. *L. lactis* products (potentially) useful in foods and in health care

<table>
<thead>
<tr>
<th>Products</th>
<th>Applications/functions</th>
<th>Type</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteriocin</td>
<td>Antimicrobial, preservative</td>
<td>Peptides or Proteins</td>
<td>[16]</td>
</tr>
<tr>
<td>Casocidin</td>
<td>Antimicrobial, preservative</td>
<td></td>
<td>[17]</td>
</tr>
<tr>
<td>Isracidin</td>
<td>Antimicrobial, preservative</td>
<td></td>
<td>[17]</td>
</tr>
<tr>
<td>Kappacin</td>
<td>Antimicrobial, preservative</td>
<td></td>
<td>[17]</td>
</tr>
<tr>
<td>Brazzein</td>
<td>Sweetener</td>
<td></td>
<td>[18]</td>
</tr>
<tr>
<td>Mabinlin II</td>
<td>Sweetener</td>
<td></td>
<td>[19]</td>
</tr>
<tr>
<td>L-alanine</td>
<td>Sweetener</td>
<td></td>
<td>[20]</td>
</tr>
<tr>
<td>Diacetyl</td>
<td>Sweetener</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td>Acetoin</td>
<td>Flavoring</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td>Linalool</td>
<td>Flavoring</td>
<td></td>
<td>[22]</td>
</tr>
<tr>
<td>Branched chain aldehydes</td>
<td>Flavoring</td>
<td></td>
<td>[23]</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>Flavoring, preservative</td>
<td></td>
<td>[24]</td>
</tr>
<tr>
<td>Gamma amino butyric acid (GABA)</td>
<td>Medical</td>
<td></td>
<td>[25]</td>
</tr>
<tr>
<td>Hyaluronic acid (HA)</td>
<td>Cosmetic, medical</td>
<td></td>
<td>[26]</td>
</tr>
<tr>
<td>Folate (Vitamin B9)</td>
<td>Health supplements</td>
<td></td>
<td>[27]</td>
</tr>
<tr>
<td>Riboflavin (Vitamin B2)</td>
<td>Health supplements</td>
<td></td>
<td>[28]</td>
</tr>
<tr>
<td>Menaquinone (Vitamin K2)</td>
<td>Health supplements</td>
<td></td>
<td>[29]</td>
</tr>
</tbody>
</table>
Transcriptional regulation in *L. lactis*

The regulation of gene expression (GE) has been a key research topic for decades, as it is highly important for the survival of bacteria in ever-changing environments. For LAB, most studies to date have focused on the regulation that takes place at the mRNA or protein levels. Transcriptional regulation can be achieved by modulating the initiation of gene transcription with sigma factors or by employing regulator proteins targeting specific DNA motifs and elements and in this way can act as activators, enhancers or repressors [36]-[38]. Of the many GE systems discovered in *L. lactis*, the transposon-based two-component system regulating the biosynthesis, immunity and secretion of the bacteriocin nisin is arguably the most important. It has been redesigned into a versatile and widely used tool for the deliberate (over) expression of a wealth of genes in *L. lactis* and a number of other LAB [39], [40]. This will be further detailed in the next section.

The nisin-controlled expression (NICE) system

The NICE system comprises a nisin-controlled two-component signal transduction system (NisRK) that regulates the expression of a gene of interest (*goi*) [41] (see Figure 2). The inducer molecule nisin is a membrane-permeating antimicrobial peptide that can be sensed by the membrane-embedded protein NisK, a histidine kinase. In the presence of nisin, NisK autophosphorylates, upon which it phosphorylates (P$_i$) the transcription regulator NisR. Phosphorylated NisR subsequently activates the promoter P$_{nisA}$ upstream of the *goi* to facilitate *goi* transcription. The *nisRK* genes were integrated into the chromosome of the *L. lactis* strain MG1363 under the control of a constitutive promoter, resulting in the strain NZ9000 (see Figure 2). The nisin-controllable expression (NICE) platform was built using the pNZ-series of vectors, which include a chloramphenicol resistance gene (CmR) and in which the *goi* is placed downstream of the nisin-inducible promoter P$_{nisA}$[42] (Figure 2). The expression level of the *goi* can be controlled by modulating the (non-toxic) amount of nisin added to the cells. This system is generally used for gene–protein overexpression in *L. lactis*. 
Figure 2 The nisin controlled gene expression system (NICE) in *L. lactis*. The antimicrobial peptide nisin activates NisK to phosphorylate NisR. Phosphorylated NisR functions as a transcriptional activator of the promoter $P_{nisA}$. The genes of interest (gois) are located downstream of the nisin-controlled promoter $P_{nisA}$. Purple form: ribosome. Orange beads: the proteins of interest (poi).

**Post-transcriptional regulation**

During the past several decades, post-transcriptional regulation has been revealed as an important mechanism of rapid control of gene expression in bacteria. The rather weak correlation between RNA and protein abundance has pinpointed the existence and important role of post-transcriptional regulation: the control of gene expression occurring after transcription but before translation [43]. Different post-transcriptional regulators have been identified and divided into two types: small RNAs (sRNAs) and RNA-binding proteins (RBPs). The general mechanisms of post-transcriptional gene regulation mainly include: adjustment of mRNA decay by recruitment or prevention of ribonuclease access; influencing translation initiation efficiency by blocking or releasing the blockage of the ribosome binding site (RBS); and controlling transcript elongation by hindering RNA polymerase progression.
RNA-binding proteins

The post-transcriptionally active regulatory RNA-binding proteins (RBPs) can exert regulation by influencing the stability, translation initiation accessibility, or transcript elongation of their target RNAs. These targets can be either mRNAs or sRNAs. Regulation by RBPs can occur in three different ways: (i) regulating the efficiency of translation initiation by modulating RBS accessibility to RNases, either by directly targeting the mRNA or indirectly by interacting with regulatory molecules such as sRNAs; (ii) affecting RNA stability and degradation by regulating the binding of other RBPs, mainly RNases, to RNA elements, with or without changing the target RNA secondary structure; and (iii) modulating transcription elongation through the involvement of terminator or anti-terminator structure formation. A schematic overview of these regulatory mechanisms is presented in Figure 3.

Translation initiation involves a broad region from -20 to +19 surrounding the start codon on mRNAs that acts as the interaction region of a ribosome initiating translation [44]. Some RBPs directly compete with ribosomes by blocking the RBS of a gene on an mRNA [45] [46] (Figure 3A). Some sRNA-mediated chaperone RBPs in *E. coli* can recruit the degradosome. This RNA-degrading protein complex usually consists of RNase E, which also functions as a scaffolding protein, and several other components, such as polynucleotide phosphorylase (PNPase), an exoribonuclease that catalyzes the processive phosphorolytic degradation of RNA from the 3’ end in various bacteria [47]; enolase, a glycolytic enzyme that interconverts 2-phosphoglycerate and phosphoenolpyruvate [48]; and ATP-dependent RNA helicase B (RhlB) [49]. Together, they accomplish the degradation of the mRNA target [50], [51] (Figure 3B).

In addition, RBSs can be freed or sequestered by RBPs, causing a secondary structural change that either activates or represses the initiation of translation, respectively [52]-[55] (Figure 3C).

Ribonucleases in *L. lactis*

Regulation of the stability of RNAs is also an important mechanism in gene expression control, as it can affect the number of transcripts that can be sequentially translated or the amount of sRNA that can perform specific regulatory functions. RNA stability can be directly affected by ribonucleases. The primary or secondary sequence of an RNA can recruit ribonucleases. *E. coli* single-strand RNases (e.g., RNase E, RNase G, and RNase Y) or double-strand RNases (e.g., RNase III) can be involved in mRNA decay. In *E. coli*, mRNA breakdown usually begins with (an) endoribonucleolytic cleavage(s), followed by the action of a combination of endo- and exonucleases such as PNPase and RNase II but rarely RNase R [56]. The two major and typical enzymes involved in mRNA degradation in *E. coli* are endonuclease RNase E and 3’-exonuclease RNase II. These enzymes are shown as model ribonucleases in
**Figure 3.** Homologs of these two ribonucleases are absent in *L. lactis*. Instead, the major decay-initiating endonucleases in *L. lactis* are most likely analogous to those in *B. subtilis*, RNase Y and PNPase. In addition, *L. lactis* contains two putative RNase R homologs. The specific roles of ribonucleases in bacterial RNA metabolism have been reviewed previously [57]. Table 2 shows a summary comparing the ribonucleases in *E. coli* and *B. subtilis* as well as the homologs of the various RNases in *L. lactis*.

RBPs can influence mRNA stability indirectly by competing with RNases (Figure 3D) to increase RNA stability [82] or by recruiting RNases to promote RNA decay (Figure 3E). The stability of RNA can also be positively or negatively affected by RBP-mediated changes in RNA secondary structure that lead to the disclosure or burial of RNase target sites (Figure 3F). Additionally, *E. coli* PAPI (poly A polymerase I) can add a single-strand poly(A) extension at the 3’-ends of transcripts, which promotes their degradation [83] (Figure 3E).

Another mechanism of post-transcription regulation by RBPs is the promotion or interruption of transcription elongation. Two types of terminators exist in bacteria: those that are dependent and those that are independent of the small helicase protein Rho. Rho-independent terminators generate a self-annealing hairpin structure, usually located at the end of the gene/operon transcript. Premature transcription termination can also occur in the 5’-region upstream of the RBS or within the gene body. Both will exclude transcript elongation to full-length mRNA. These terminator structures can be stabilized by RBPs (Figure 3G). Alternatively, an anti-terminator can be formed by recruiting RBPs. This stops termination at specific sites by preventing the formation of adjacent terminator (Figure 3H). Usually, these two structures are mutually exclusive and quite specific for the regulation of expression of only a few genes [84]. Rho-dependent terminators require Rho factor to recognize an unstructured cytosine-rich sequence known as the Rho utilization site (*rut*) to disrupt transcription [85]. RNA-binding proteins can expose the *rut* site and allow rho-dependent termination to occur (Figure 3I).
### Table 2. Summary of main ribonucleases in *E. coli* and *B. subtilis*, and the homologs in *L. lactis*

<table>
<thead>
<tr>
<th>E. coli</th>
<th>B. subtilis</th>
<th>Gene</th>
<th>Activity type</th>
<th>Major function</th>
<th>Ref</th>
<th>Homolog in L. lactis</th>
<th>Gene alias</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNase I</td>
<td></td>
<td>rna</td>
<td>Endonuclease</td>
<td>Clearing of cellular RNA no longer required</td>
<td>[58]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNase P</td>
<td>RNase P</td>
<td>rnpA/</td>
<td>Endonuclease</td>
<td>5' tRNA maturation</td>
<td>[59]</td>
<td>Yes</td>
<td>ilmg_0142</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rnpB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNase E</td>
<td></td>
<td>rne</td>
<td>Endonuclease</td>
<td>mRNA decay, rRNA and tRNA maturation</td>
<td>[60]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNase G</td>
<td></td>
<td>rng</td>
<td>Endonuclease</td>
<td>rRNA maturation</td>
<td>[61]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YbeY</td>
<td>YbeY</td>
<td>ybeY</td>
<td>Endonuclease</td>
<td>rRNA maturation and quality control</td>
<td>[62]</td>
<td>Yes</td>
<td>ilmg_1487</td>
</tr>
<tr>
<td>RNase Y</td>
<td></td>
<td>rny</td>
<td>Endonuclease</td>
<td>mRNA decay</td>
<td>[63]</td>
<td>Yes</td>
<td>ilmg_2156</td>
</tr>
<tr>
<td>RNase Z</td>
<td></td>
<td>rnz</td>
<td>Endonuclease</td>
<td>tRNA 3' maturation</td>
<td>[64]</td>
<td>Yes</td>
<td>ilmg_0604</td>
</tr>
<tr>
<td>RNase M5</td>
<td></td>
<td>rnmV</td>
<td>Endonuclease</td>
<td>5S rRNA maturation</td>
<td>[65]</td>
<td>Yes*</td>
<td>ilmg_1883</td>
</tr>
<tr>
<td>RNases III</td>
<td>RNases III</td>
<td>mc</td>
<td>Endonuclease, double-strand specific</td>
<td>rRNA maturation</td>
<td>[66]</td>
<td>Yes</td>
<td>ilmg_1753</td>
</tr>
<tr>
<td></td>
<td>Mini-RNases III</td>
<td>mrc</td>
<td>Endonuclease, double-strand specific</td>
<td>23S rRNA maturation</td>
<td>[67]</td>
<td>Yes*</td>
<td>ilmg_2039</td>
</tr>
<tr>
<td>RNase HI</td>
<td></td>
<td>rnhA</td>
<td>Endonuclease on RNA primers and R-loops in DNA</td>
<td>DNA replication</td>
<td>[68]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNase HII</td>
<td>RNase HII</td>
<td>rnhB</td>
<td>Endonuclease on the RNA hybridized to DNA</td>
<td>removal of ribonucleotides in DNA</td>
<td>[69]</td>
<td>Yes</td>
<td>ilmg_1176</td>
</tr>
<tr>
<td>RNase HIII</td>
<td>RNase HIII</td>
<td>rnhC</td>
<td>Endonuclease on the RNA hybridized to DNA</td>
<td>Removal of ribonucleotides in DNA</td>
<td>[70]</td>
<td>Yes</td>
<td>ilmg_2549</td>
</tr>
<tr>
<td>RNase LS</td>
<td></td>
<td>rnlA</td>
<td>Endonuclease</td>
<td>mRNA decay</td>
<td>[71]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNase BN</td>
<td></td>
<td>rbn</td>
<td>Endonuclease/3' to 5' exonuclease</td>
<td>sRNA decay, mRNA stabilization</td>
<td>[72]</td>
<td>Yes*</td>
<td>ilmg_0578</td>
</tr>
<tr>
<td>Rnase J1</td>
<td></td>
<td>rnjA</td>
<td>5' to 3' exonuclease/endonuclease</td>
<td>16S rRNA 5' maturation, mRNA decay</td>
<td>[73]</td>
<td>Yes*</td>
<td>ilmg_0876</td>
</tr>
<tr>
<td>Rnase J2</td>
<td></td>
<td>rnjB</td>
<td>5' to 3' exonuclease/endonuclease</td>
<td>16S rRNA 5' maturation, mRNA decay</td>
<td>[73]</td>
<td>Yes*</td>
<td>ilmg_0302</td>
</tr>
<tr>
<td>Rnase AM</td>
<td></td>
<td>Trp/Y</td>
<td>5' to 3' exonuclease</td>
<td>mRNA decay</td>
<td>[74]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rnase II</td>
<td></td>
<td>rnb</td>
<td>3' to 5' exonuclease</td>
<td>mRNA decay, rRNA and tRNA maturation</td>
<td>[75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNPase</td>
<td>PNPase</td>
<td>pnp/pnpA</td>
<td>3' to 5' exonuclease</td>
<td>mRNA decay, rRNA and tRNA maturation</td>
<td>[76]</td>
<td>Yes</td>
<td>ilmg_2044</td>
</tr>
<tr>
<td>Rnase D</td>
<td></td>
<td>rnd</td>
<td>3' to 5' exonuclease</td>
<td>tRNA maturation</td>
<td>[77]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rnase T</td>
<td></td>
<td>rnt</td>
<td>3' to 5' exonuclease</td>
<td>tRNA and rRNA maturation</td>
<td>[78]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rnase R</td>
<td>Rnase R</td>
<td>rnr</td>
<td>3' to 5' exonuclease</td>
<td>mRNA decay, rRNA and tRNA maturation, rRNA quality control</td>
<td>[79]</td>
<td>Yes*</td>
<td>ilmg_1586, ilmg_1304</td>
</tr>
<tr>
<td>Rnase PH</td>
<td>Rnase PH</td>
<td>rph</td>
<td>3' to 5' exonuclease</td>
<td>tRNA maturation, rRNA and tRNA degradation</td>
<td>[80]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orn</td>
<td></td>
<td>orn</td>
<td>3' to 5' exonuclease</td>
<td>Removal of short oligonucleotides</td>
<td>[81]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Putative homolog, not verified in literature
Figure 3. Overview of RNA-binding protein regulatory mechanisms in bacteria. (A-C) The mechanisms by which RBPs regulate translation initiation are shown. (A), Direct blockade of the ribosomal binding site (RBS) to repress translation. (B) sRNAs targeting RBS via chaperone proteins recruit the degradosome, causing mRNA decay. (C) Changes in the secondary structure surrounding the RBS affect the accessibility of RNase E. (D-F) RBP regulation is involved in RNA degradation. Stability of both mRNA and sRNA can be regulated. RBS and protein coding sequence (CDS) are represented by the dotted lines in (D) and (F), respectively. (D), Directly blocking RNase recognition sites represses decay. (E) Recruiting RNases to promote decay. (F) Effect of changing the accessibility of RNase E recognition sites, whether buried or exposed in the secondary structure. (G-I) Changes in the RNA secondary structure caused by RBPs modulate transcription elongation. (G), Stabilizing the terminator structure. (H) Stabilizing the antiterminator structure. (I), Making changes in secondary structure, releasing rho utilization (rut) site to promote rho-dependent termination. (For all panels: see text for further details)
Post-transcriptional RNA-binding proteins in *L. lactis*

In addition to ribonucleases, other RBPs with specific and unique functions have been revealed recently. The CsrA/RsmA family of proteins comprises typical RBPs that can bind to specific GGA motifs in RNA in many Gram-negative bacteria [86]. RsmA/CsrA proteins can post-transcriptionally affect the stability of mRNAs and transcript elongation by binding to RNA secondary structures in 5’-UTRs; they can also repress translation. In these ways they have been shown to impact the regulation of biofilm formation, stress responses, virulence and carbon metabolism [87]–[90]. Cold-shock proteins (CSPs) are also well-characterized RBPs. CSP homologs are present in 2 to up to 9 copies in Gram-positive and Gram-negative bacterial species [91]. Some CSPs act as global regulators of post-transcription regulation by modulating RNA secondary structure to adjust the accessibility of RBSs or RNases or affecting transcription termination. For instance, overproduction of the major *E. coli* CSP CspA, as well as its homologous RNA chaperones CspE and CspC, can increase the expression of a cold shock induced operon consisting of the genes *nusA, infB, rbfA* and *pnp* through a transcription antitermination mechanism [92]. CspB in *B. subtilis* has also been demonstrated to have mRNA-binding capacity [93]. In *L. lactis*, at least five CSP-encoding genes, including two tandem sets (*cspA/cspB* and *cspC/cspD*), have been identified on the chromosome of *L. lactis* MG1363. Both gene sets are cold inducible, while the single *cspE* gene is constitutively expressed [90]. Although CSPs contain 2 highly conserved RNA-binding motifs, RNP-1 and RNP-2 [95], it has not yet been reported whether these 5 CSPs function as RNA chaperones in *L. lactis*. Table 3 lists the best known post-transcriptionally functioning RBPs in bacteria. Although a large number of RBPs have been identified in various bacterial species, and some of those seem to be widely conserved, many of these RBPs are absent in *L. lactis*. Most notably is the absence of homologs of the sRNA chaperones Hfq and ProQ (see below). To date, only a very limited number of specific RBPs have been identified in *L. lactis*. One of them, the BglG family member BglR, shares 30% identity with BglG from *E. coli*. BglG is involved in β-glucoside utilization in *E. coli*, functioning as a transcriptional antiterminator of the operon *bglGFBH* in this organism [94]. In the absence of β-glucosides, operon transcription is terminated at a region immediately upstream of the *bglG* gene, while BglG can bind to a specific sequence that partly overlaps the terminator in the presence of β-glucosides. This prevents termination and leads to full operon expression and β-glucoside consumption. The protein BglR in *L. lactis* is also involved in β-glucoside utilization and is functionally equivalent to its homolog in *E. coli* [94]. Upstream of *L. lactis* bglR, a putative transcription terminator was identified that is highly homologous to the RNA-binding site recognized by *E. coli* BglG. The other reported *L. lactis* RBP is PyrR. Its gene is the first in the *pyrRPB-carA* operon responsible
for pyrimidine biosynthesis. In the middle of the pyr transcript, two mutually exclusive structures can be folded, which result in antitermination or termination. Based on its high similarity to the binding sequence known for B. subtilis PyR, a binding site for L. lactis PyrR (5’-UCCAGAGGCAUGCAAG-3’) was proposed to be present in the stem–loop structure [96].

<table>
<thead>
<tr>
<th>RBP</th>
<th>Species</th>
<th>Post-transcriptional</th>
<th>Interaction</th>
<th>Ref</th>
<th>Homolog</th>
</tr>
</thead>
<tbody>
<tr>
<td>RsmE</td>
<td>Pseudomonas fluorescens</td>
<td>Regulation of translation</td>
<td>Yes</td>
<td>[97]</td>
<td></td>
</tr>
<tr>
<td>RsmN/F</td>
<td>Pseudomonas fluorescens</td>
<td></td>
<td>Yes</td>
<td>[98]</td>
<td></td>
</tr>
<tr>
<td>BpuR</td>
<td>Borrelia burgdorferi</td>
<td></td>
<td></td>
<td>[99]</td>
<td></td>
</tr>
<tr>
<td>ThrRS</td>
<td>Escherichia coli</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hfq</td>
<td>Escherichia coli</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CsrA/RsmA</td>
<td>Escherichia coli, Pseudomonas</td>
<td></td>
<td>Yes</td>
<td>[101]</td>
<td></td>
</tr>
<tr>
<td>CsrD</td>
<td>Escherichia coli</td>
<td></td>
<td>Yes</td>
<td>[102]</td>
<td></td>
</tr>
<tr>
<td>ProQ</td>
<td>Escherichia coli</td>
<td></td>
<td>Yes</td>
<td>[103]</td>
<td></td>
</tr>
<tr>
<td>RapZ</td>
<td>Escherichia coli</td>
<td></td>
<td>Yes</td>
<td>[104]</td>
<td></td>
</tr>
<tr>
<td>RhlB</td>
<td>Escherichia coli</td>
<td></td>
<td>Yes</td>
<td>[105]</td>
<td>Yes</td>
</tr>
<tr>
<td>RNaseE</td>
<td>Widely conserved</td>
<td></td>
<td></td>
<td>[106]</td>
<td></td>
</tr>
<tr>
<td>YbeY</td>
<td>Widely conserved</td>
<td></td>
<td>Yes</td>
<td>[107]</td>
<td>Yes</td>
</tr>
<tr>
<td>flbT</td>
<td>Caulobacter crescentus</td>
<td></td>
<td></td>
<td>[108]</td>
<td></td>
</tr>
<tr>
<td>BglG</td>
<td>Escherichia coli</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EutV</td>
<td>Enterococcus faecalis</td>
<td></td>
<td></td>
<td>[109]</td>
<td></td>
</tr>
<tr>
<td>NusA</td>
<td>Bacillus subtilis</td>
<td></td>
<td></td>
<td>[110]</td>
<td></td>
</tr>
<tr>
<td>PyrR</td>
<td>Bacillus subtilis</td>
<td></td>
<td></td>
<td>[111]</td>
<td></td>
</tr>
<tr>
<td>HutP</td>
<td>Bacillus subtilis</td>
<td></td>
<td></td>
<td>[112]</td>
<td>Yes</td>
</tr>
<tr>
<td>TRAP</td>
<td>Bacillus subtilis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSPs</td>
<td>Widely conserved</td>
<td>Global regulation of translation, RNA degradation and termination</td>
<td>Yes</td>
<td>[115]</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The proteins RapZ and YbeY have been relatively recently characterized in E. coli as having RNA-binding activity. Homologs of both are present in L. lactis. YbeY is an endoribonuclease involved in 16S rRNA 3’-end processing in a number of bacterial species. L. lactis YbeY seems to be essential as several attempts to mutate or delete this gene were unsuccessful (this thesis). RapZ is an sRNA-binding protein that regulates amino sugar metabolism in E. coli. L. lactis RapZ (Lmg_1557 of strain MG1363) shares 59.5% identity with RapZ of E. coli. Both YbeY and RapZ will be further detailed later in this Introduction.

An approach based on the acidic guanidinium thiocyanate-phenol chloroform (AGPC) phase partition, termed orthogonal organic phase separation (OOPS) [116], has recently been
successfully used for mining the landscape of RNA-binding proteins in *L. lactis* MG1363 (Eduardo Hernandez Ortega, MSc. pers. comm.), which could inspire RBP research and studies on sRNAs in *L. lactis* in the near future.

Small regulatory RNAs in bacteria

Bacterial small regulatory RNAs (sRNAs) mostly range in size from 50 to 300 nucleotides; they are usually noncoding and have been identified as important regulators in bacteria during the last two decades. sRNA genes were first identified in plasmids, transposons and phages [117], [118]. The first chromosomally derived sRNA, *E. coli* MicF, was shown to inhibit translation by blocking the RBS of the gene encoding the major outer membrane porin OmpF [119]. Hundreds of new or candidate sRNAs have been identified since the start of the 21st century by major advances in experimental approaches, including cloning-based techniques, tiled microarrays with full-genome coverage, multilayered computational searches and deep DNA/RNA sequencing [120]. sRNAs have been implicated in a host of important biological processes, including stress response regulation, ion homeostasis, quorum sensing, and biofilm formation [121]–[123].

Regulatory mechanisms of sRNAs

Genes of regulatory sRNA share the common characteristics of being typically located within intergenic regions, having independent promoters and containing Rho-independent terminators. These features were used to identify new sRNAs by employing systematic computational tools [124]. sRNAs can be classified into two types: those base-pairing with and regulating target mRNAs, and those that regulate through the recruitment via specific RNA structures of functional proteins. The majority of sRNAs belong to the first type and act by base-pairing with mRNAs. Several of these sRNAs require the aid of so-called sRNA chaperones (see below). These sRNAs can be divided into trans-encoded sRNAs and cis-encoded sRNAs (Figure 4A). Trans-encoded sRNAs are usually located intergenically and operate through imperfect base-pairing with their target mRNA(s). Most sRNAs discovered to date are trans-encoded sRNAs. Regulation in trans [125] is usually negative but base-pairing sRNAs can also act positively with various regulatory effects, such as affecting translation or RNA stability. Base-pairing sRNAs can block RNase recognition sites or increase their translation or, in contrast, inhibit translation or promote RNA degradation (Figure 4B). Dissimilar to trans-encoded sRNAs, cis-encoded sRNAs, also labeled antisense RNAs (asRNAs), are regulated by full complementary base-pairing with their targets. Usually, the asRNA is located on the DNA strand opposite to that of the coding gene which it controls. The majority of cis-encoded sRNAs are expressed from bacteriophage genomes, transposons or plasmids.
The latter usually function to appropriately control the plasmid copy number [126]. An asRNA usually results in target degradation. Some cis-encoded sRNAs can also bind to a target region between two genes in an operon. This interaction can cause transcription termination of the upstream gene or attract RNases to the target region, causing mRNA degradation [127].

**Figure 4. sRNA regulatory mechanisms in bacteria.** (A) Intergenically transcribed sRNA and asRNAs transcribed from the strand opposite that of the target gene. (B) Positive or negative regulatory activity of sRNAs via modulation of RNase activity or translation initiation.

**Riboswitches**

Riboswitches are another type of RNA regulator; they are located in the 5′-UTRs of many genes. These RNA structures act in cis. By sensing ions, temperature, or small metabolites [128], [129], riboswitches shift their RNA structure conformation. They can regulate at the transcriptional or post-transcriptional level (e.g., transcription termination, translation initiation, and transcript stability) [130], [131]. Although they function in different ways, there is no absolute boundary between sRNAs and riboswitches. For example, the riboswitch-containing sRNA EutX can sequester the response regulator EutV to control gene expression by sensing the absence of adenosyl cobalamin in *E. faecalis* [132].
Synthetic sRNAs have been designed to target and affect gene expression for metabolic engineering purposes. An sRNA was designed to target riboswitches to promote the antitermination folding, thus preventing transcription termination and increasing riboswitch gene ribB expression in *B. subtilis* to engineer an improved riboflavin producer strain [133].

**Proteins participating in the base pairing of sRNAs**

-Chaperone proteins

As sRNA-mRNA interactions are in many cases based on imperfect base pairing (see above), chaperone proteins are often required to facilitate and improve the accuracy of the binding of the two RNAs. The most prominent RNA chaperone proteins in many bacterial species are Hfq and ProQ [134]. Hfq is a highly abundant, hexameric small protein found in many bacteria. A large number of bacterial sRNAs, particularly *trans*-encoded sRNAs, are reported to be Hfq dependent. For instance, approximately 30% of sRNAs in *E. coli* rely on Hfq for proper functioning [135]. Usually, Hfq acts by disrupting the secondary structures of target RNAs, remodeling them and exposing the bases for pairing to their complementary partners on the sRNA. Hfq seems to have a binding preference for (ARN)n motifs (A, adenine; R, guanine/adenine; N, adenine/guanine/cytosine/thymine), which are frequently found in mRNAs and the uridine-rich 3'-end of sRNAs [136], [137]. Hfq does not only interact with sRNAs but can also bind to rRNA, tRNA and even DNA, suggesting that it has other functions [138]–[140]. ProQ is a recently identified RNA chaperone that belongs to the FinO family of bacterial RNA chaperones [141]. It is a monomeric small protein consisting of two domains: an N-terminal domain that is similar to the RNA-binding FinO domain that is present in the RNA chaperones ProQ of *S. enterica* and RocC of *Legionella pneumophila* and a C-terminal domain that is critical for RNA strand exchange activity [142]. The surface of the FinO domain of ProQ contains positively and negatively charged residues, which could lead to RNA binding via electrostatic interactions or hydrogen bonding. The ProQ chaperone has a preference for structure rather than nucleotide sequence in target RNA via FinO. The purified N-terminal FinO domain was shown to bind double-stranded (ds) RNAs in *E. coli* [143]–[145]. In contrast to Hfq, which particularly promotes the action of *trans*-encoded sRNAs, ProQ-assisted sRNAs mostly act in *cis*. Thus, ProQ promotes the perfect base-pairing of asRNAs with their targets, but it can also assist in imperfect base-pairing between target and *trans*-encoded sRNAs [104]. In addition to Hfq and ProQ, CsrA was reported to be a third sRNA chaperone in *E. coli*. It promotes the base-pairing of the *trans*-encoded sRNA SR1 with its target mRNA *ahrC*. This interaction causes a change in the RBS conformation in such a way that it prevents *ahrC* translation [146].
-Ribonucleases

RNase E cleaves single-stranded (ss) RNA. It prefers A/U-enriched regions with an upstream stem–loop structure. It can also be involved in the cleavage of sRNA-mRNA duplexes. The Hfq-mediated hybrid between sRNA CpxQ and mRNA cpxP in *Salmonella enterica* is cleaved at the duplex region through a specific 3′-UTR-exonucleolytic process involving RNase E [147].

PNPase is the main enzyme that causes the degradation of sRNAs that do not interact with Hfq [101]. For example, the degradation of the sRNAs RybB and asRNA MicA in *E. coli* is controlled by PNPase. The enzyme can also degrade sRNAs in the absence of their mRNA targets [148].

RNase III is known mainly as an endoribonuclease involved in the processing of dsRNAs, for instance during rRNA processing. RNase III-mediated cleavage of sRNA-mRNA duplexes is mostly reported in bacteria. For instance, RNase III degrades the dsRNA duplex of the sRNA MicA and its target mRNA *ompA* in *S. typhimurium* [149]. RNase III also promotes 3′-UTR-derived sRNA processing. The sRNA RsaC can be released by RNase III from the 3′-UTR of *mntABC* mRNA in *Staphylococcus aureus* [150].

In addition to the three most studied ribonucleases mentioned above, the ribonucleases YbeY and RNase R were also proven to participate in the regulation of sRNAs. YbeY was reported to be involved in sRNA-dependent regulation in *Sinorhizobium meliloti*, *Vibrio cholerae* and *E. coli* [151]–[153]. RNase R, a unique 3′-5′ exoribonuclease that degrades highly structured RNAs, has also been implicated in the regulation of sRNAs in many bacteria [154], [155].

-Sigma factor

Sigma (σ) factors play a prominent role in transcriptional regulation. All bacterial genomes encode at least one essential σ factor responsible for the transcription of housekeeping genes, while many also encode genes for specific σ factors; these genes respond to specific stimuli, enabling a shift to other transcription repertoires. Seven sigma factors, the housekeeping σ-factor σ^{70} (*rpoD*) and σ^{19}, σ^{24}, σ^{28}, σ^{32}, σ^{38} and σ^{54} (*fcl*, *rpoE*, *rpoF*, *rpoH*, *rpoS* and *rpoN*, respectively), have been described in *E. coli*. Although sRNAs mainly act post-transcriptionally (see above), certain sRNAs affect transcription through their effect on sigma factors. For instance, the sRNAs DsrA, ArcZ and RprA of *E. coli* can bind the 5′-UTR of the *rpoS* gene, which encodes the general stress response sigma factor σ^{38}, to repress Rho-dependent transcription termination [156]. The expression of sRNA genes can also be controlled by sigma factors. The *S. enterica* sRNAs RybB and MicA, for instance, are slightly expressed in the exponential phase of growth but more so in the stationary phase, under the control of σ^{6} [157]. As for sigma
factors in *L. lactis* only 1 has been described to date, the equivalent of the housekeeping σ-factor σ^{70} (rpoD) Genome Sequences of *Lactococcus lactis* MG1363 [158].

**sRNAs binding to proteins**

Most of sRNAs interact with mRNAs, only a few protein-binding sRNAs have been identified and characterized. Among them, CsrB, 6S and GlmY are the most well studied. CsrB and GlmY regulate at a post-transcriptional level, while 6S operates at the level of DNA. They act by mimicking the target motif or structure of the cognate protein or sRNA (Figure 5).

The sRNAs CsrB and CsrC can modulate the activity of the RBP CsrA, a carbon storage regulator protein involved in the regulation of carbon utilization in *E. coli*. The binding site of CsrA is the GGA sequence (the seed) located in the 5'-UTR of the target mRNA. By binding to the sRNA CsrB or CsrC, CsrA can either repress or activate mRNA translation. Since the sRNAs CsrB and CsrC contain 22 and 13 GGA motifs, respectively, they can sequester the CsrA protein quite effectively, resulting in the stability and translation of the target transcript(s) (Figure 5A). Homologs of the CsrB family of sRNAs are present in a wide range of bacteria, including *Salmonella, Pseudomonas, Erwinia carotovora* and *Vibrio* species. The (putative) sequestering mechanism in these species would be operative in regulating various biological processes, such as cell motility, quorum sensing, biofilm formation, and pathogenesis [159]. In addition to its role as a sequestering protein, *B. subtilis* CsrA is involved in complex formation between sRNAs and mRNA targets and, thus, also functions as an RNA chaperone similar to Hfq and ProQ [146].

The sRNA 6S acts as a global regulator, downregulating transcription, especially in the stationary phase, by binding to the σ^{70} component of the *E. coli* RNA polymerase holoenzyme. By mimicking the open promoter structure (Figure 5B), 6S represses σ^{70}-regulated transcription and promotes the transcription of some σ^{S}-specific promoters [160], [161]. 6S was first discovered in *E. coli*, after which it was shown to be widely conserved in other bacterial species. The 6S RNAs share a common secondary structure but have little primary sequence similarity [162]. Some bacterial species have multiple 6S RNAs with independent functions [163]. The expression pattern of 6S in different species are quite divergent, indicating that 6S can have both common and species-specific functions. The presence of the 6S sRNA in *L. lactis* has also been reported recently; it was suggested that the expression of *L. lactis* 6S is dependent on the carbon source (see below) [164].

An interesting sRNA/protein interaction has recently been uncovered in *E. coli*, exemplifying the complexity of the control mechanisms underlying some sRNAs. The sRNA GlmY regulates glucosamine-6-phosphate (GlcN6P) levels in *E. coli* together with GlmZ, an sRNA that stimulates
the translation of the enzyme glucosamine-6-phosphate synthase (GlmS) [142]. GlmZ and GlmY share a high degree of sequence identity (63%) and have similar predicted three-dimensional structures [165]. Thus, both of the sRNAs can sequester their target, the RapZ protein, via a mimicry mechanism. When the GlcN6P level is high, GlmY is present in low amounts. In that case, GlmZ will bind to RapZ, which attracts the RNA degradosome, leading to the degradation of GlmZ. As GlmZ activates the expression of \textit{glmS}, the degradation of GlmZ leads to a block in GlmS synthesis, and the level of GlcN6P will decrease. When the level of GlcN6P is low, GlmY will sequester RapZ and GlmZ is now free to base-pair with the \textit{glmS} transcript. Together with chaperone protein Hfq, GlmZ activates the synthesis of GlmS, leading to GlcN6P production (Figure 5C).
Figure 5. Regulatory mechanisms of protein-binding sRNAs. (A) Increased sRNA CsrB production sequesters the CsrA protein through the GGA hairpins to which CsrA binds. (B) sRNA 6S sequestering of σ70 reduces transcription of certain housekeeping genes. (C) Two *E. coli* sRNAs, GlmY and GlmZ, regulate the production of glucosamine-6-phosphate synthase (GlmS) by competing with each other for binding to RapZ (Figure adapted from Gopel et al. [142]).
Chapter 1

Dual-function sRNAs

Although sRNAs are usually noncoding, some have a dual function and specify a peptide or protein. The first identified dual-function sRNA was S. aureus RNAIII, one of the largest regulatory sRNAs (514 nucleotides). It primarily acts as an asRNA, pairing with mRNAs encoding virulence factors. It also contains an ORF specifying the 26-amino acid cytotoxic peptide δ-hemolysin (hld), which targets host cell membranes and causes cell lysis [166]. The only known dual-function sRNA to date in Gram-negative bacteria is SgrS; it can block the translation of the transcript of the ptsG gene, which encodes a sugar-phosphate transporter in E. coli. A conserved ORF within the SgrS RNA sequence, called sgrT, can be translated to the 43-amino-acid SgrT protein under conditions of glucose-phosphate stress in E. coli [167]. The discovery of dual-function sRNAs has led to the interesting assumption that some mRNAs encoding small proteins might have additional functions as sRNA regulators [119].

Discovery and function of sRNAs in L. lactis

Most studies on sRNA-mediated regulatory mechanisms are performed in E. coli. Recently, the sRNAs of Gram-positive model bacteria have drawn increased attention. The sRNA landscape in L. lactis has been revealed in the last decade by employing differential RNA sequencing [168]. Sixty possible asRNAs and 186 trans-encoded putative sRNAs were identified and annotated in the L. lactis paradigm strain MG1363. A large number of predicted sRNAs and asRNAs are located in prophage regions, which take up about 5.5% of the L. lactis MG1363 genome [168]. Of the sRNAs, 14 were experimentally confirmed, including the abundant sRNA 6S (LLMGnc_004). A catabolite-responsive element (cre) is present upstream of the putative -35 box in the promoter of 6S. The expression of L. lactis 6S is under the control of the carbon catabolite repression protein CcpA, as has been shown by Northern hybridization of strain MG1363 growing with different carbon sources. 6S is abundant in the stationary phase and highly expressed during the exponential phase when cells are grown with galactose or cellobiose as the sole carbon source [164]. Another sRNA involved in carbon uptake and metabolism is LLMGnc_147, the gene of which also carries a possible cre site overlapping the -35 box in its promoter. LLMGnc_147 was shown by Northern blot analyses to be highly expressed in cells growing on cellobiose. The controlled expression of LLMGnc_147 during growth on cellobiose may operate via the transcriptional regulator AraC (Llmg_0962), a cellobiose-specific transcriptional activator. Overexpressing LLMGnc_147 has a beneficial effect during growth on galactose. This effect is likely caused by the increased expression of llmg_0963, which encodes a PTS IIC component that imports galactose [164].
The *L. lactis* sRNA ArgX is derived from the 3’-UTR of the gene *argR*, which specifies the transcriptional repressor ArgR [169]. ArgR can repress *arc* operon mRNA levels to regulate the arginine deiminase pathway. ArgX putatively blocks the translation of the *arcC1* transcript, which encodes a carbamate kinase. This was the first example of a transcription factor and an sRNA derived from the 3’-UTR of its transcript sharing a common target and, together, acting on the transcription and post-transcriptional levels, respectively [169].

sRNAs are key regulators of stress responses, as we have shown above; some of them were also shown to perform stress-responsive functions in *L. lactis*. The sRNAs S015 and S042 were reported to improve the acid tolerance of *L. lactis* F44 [170], [171]. The responses of sRNAs to several industrial stresses were also assessed in *L. lactis* NCDO712 [172], the parent strain of *L. lactis* MG1363. The strain NCDO712 carries 6 plasmids, some of which contribute to the properties of the strain in the dairy industry [169]. Of the 186 predicted sRNAs in *L. lactis* MG1363, 110 were differentially expressed in NCDO712 under the environmental stresses employed, suggesting that they play important roles in stress response mechanisms. sRNAs of which the expression was significantly changed under oxidative, starvation or acid stress were mostly downregulated, while under cold stress, the expression of the responding sRNAs is mostly upregulated [172].

**L. lactis** sRNA chaperones

As mentioned previously, *L. lactis* lacks homologs of the chaperones Hfq and ProQ. This indicates that chaperones are not required in *L. lactis* for the base-pairing of sRNAs and their targets or that an as yet unidentified RBP(s) acts as such a chaperone. A putative candidate chaperone in *L. lactis* MG1363 is Llmg_1487, a homolog of *E. coli* YbeY (see above). Although the main function of YbeY is its endoribonuclease activity involved in 16S rRNA 3’-end processing [174], YbeY has been reported to regulate virulence-associated sRNAs in *V. cholerae* [148]. Loss of YbeY results in effects on the expression of various sRNAs, indicating that YbeY has a potential chaperone function in modulating sRNA-mRNA interactions in both *E. coli* and *B. subtilis* [153], [154]. *L. lactis* YbeY is the object of study in Chapter 2 of this thesis.

The product of the gene *llmg_1557*, RapZ, is another candidate sRNA chaperone in *L. lactis* MG1363. The role of RapZ in the RNA-mediated regulation of amino sugar metabolism in *E. coli* has been discussed above. RapZ is conserved among most Gram-negative bacteria. *L. lactis* RapZ is conserved in Gram-positive bacteria, while only the C-terminal part of RapZ is conserved in Gram-negative and Gram-positive bacteria. This indicates that the protein, via its C-terminal, might partly be a functional equivalent to that of Gram-negative bacteria (or *E. coli*). The different N-terminal suggests that RapZ may have different function(s) in Gram-positive
bacteria than in the Gram-negatives. Structural analyses and experimental data show that the C-terminal domain of RapZ is sufficient for binding to RNA in E. coli [142]. Unpublished data (Eduardo Hernandez Ortega, MSc, pers. comm.) from Orthogonal Organic Phase Separation (OOPS) experiments in L. lactis, combined with MAPS, an RNA sequencing technology coupled with MS2-tagged sRNA and affinity purification [175], show that L. lactis RapZ indeed binds to the sRNAs CisR and ArgX in this organism.

Research on sRNAs in L. lactis is still at an early stage. Although functional characterization has been reported for only a limited number of sRNAs and most of the roles of the predicted sRNAs remain to be explored, the potential of sRNAs contributing to improving L. lactis properties for industrial purposes is promising and currently ongoing.

**Genetic tools for LAB genome engineering**

To understand the functions of sRNAs and RBPs and, indeed, the interplay of all genes and genetic elements in a (bacterial) cell, for fundamental reasons as well as for application purposes, requires being able to use sophisticated genetic engineering tools. The methodologies that have been applied in this thesis work are discussed below.

Typically, two approaches exist for generating genetic mutations in LAB, namely, random and targeted mutagenesis. The second involves recombination-mediated integration, which results in gene insertion, replacement or deletion mutations.

**Random mutagenesis systems**

Random mutagenesis has proven to be a valuable tool for investigating genes and their regulators. Generation of a genuinely random, unbiased mutant library is necessary. Two common approaches, transposon mutagenesis and insertion sequences (ISs), are used in bacteria to generate mutant libraries. However, transposon mutagenesis seems to face limitations in LAB due to low transfer and integration frequencies and a restricted number of transposon carriers. Therefore, IS-based mutagenesis is the most frequently used mutagenesis approach in LAB. For Lactobacilli IS elements can be used for the construction of suicide insertion vectors.

**Recombination-mediated genetic engineering systems**

The integration of target genes into the chromosomes of bacteria is a critical genetic tool. Homologous recombination in LAB, such as in lactobacilli, can be achieved by one plasmid or dual-plasmid systems. Systems based on temperature-sensitive integration vectors, such as the pG+host (a pWV01 derivative [176]), or nonreplicative plasmids, such as pCS1966 (used in our study), have been widely used especially in L. lactis.
Double-stranded recombin engineering

Gene deletion via double-crossover (DCO) technology based on homologous recombination has been achieved in various bacterial species, including *L. lactis* [177]-[179]. Low crossover frequencies are often a bottleneck. Various improvements have been made over the course of years to facilitate recombination and promote the efficiency and selection of the desired mutant strain [180]. Initially, DCO was achieved in *L. lactis* through the use of nonreplicating plasmids harboring selective antibiotic resistance genes. However, without a further selective marker, the isolation of the sought-after chromosomal deletion mutants occurs at quite low efficiencies. A thermosensitive plasmid was developed for the process, in which the first step involved the combination of antibiotics and growth of the strain at a nonpermissive higher temperature. The second selection step relies on a permissive lower temperature [180]. The limitation of this tool is that the nonpermissive temperature (37°C) is at the limit of growth for certain *L. lactis* strains and too high for others [180]. Another selection/counterselection tool that is currently widely used for *L. lactis* is based on the counterselection marker gene oroP [181]. In this strategy, an antibiotic resistance plasmid is used for the first crossover integration in the chromosome, while the subsequent selection of those cells in which the plasmid has been lost from the chromosome by homologous recombination is achieved by supplementing the selective plates with 5-fluoroorotate. The presence of the oroP gene on the chromosome leads to sensitivity to 5-fluoroorotate, and only those cells that have lost the plasmid construct from the chromosome survive; they are either revertants to the wild-type situation or the desired mutant [181] [182].
CRISPR and CRISPRi: RNA-mediated genetic engineering systems

Apart from the typical recombination-mediated precision genetic engineering techniques used over the last 30 years in LAB, CRISPR/Cas technologies have been developed as customizable, simple and efficient genome editing tools.

CRISPR/Cas is a prokaryotic adaptive immune system that is present in many bacteria and archaea; it acts against invading foreign DNA or RNA. It is a natural RNA-mediated, sequence-specific DNA and/or RNA editing tool that has been used to develop molecular tools for genome engineering in both prokaryotes and eukaryotes [183][184]. CRISPR/Cas systems can be categorized into six types (types I–VI) on the basis of the structure and function of their specific Cas protein [183]. The type I, II, and V systems can recognize and cleave DNA, while type VI edits RNA, and type III edits both DNA and RNA. The type-II CRISPR-Cas system constructed from Streptococcus pyogenes is especially widely characterized and used. Three distinct components are naturally required in type II CRISPR/Cas systems to fully silence foreign DNAs: the Cas9 protein, CRISPR RNA (crRNA) and trans-acting crRNA (tracrRNA). The latter two form a complex and binds to Cas9, which then acts as an RNA-guided DNA endonuclease [184]. The crRNA/tracrRNA complex can be mimicked and synthesized both in vivo and in vitro, in which a designed hairpin “handle” is coupled to a small guide RNA (sgRNA). The designed sgRNA can bind to and promote DNA cleavage by Cas9 [185] (Figure 6A). Cas9 endonuclease cleaves specifically at the sgRNA-targeted DNA sequence and induces a double-strand DNA break (DSB). The sgRNA is designed as a chimeric transcript consisting of a unique 20-nt sequence targeting the gene of interest through base-pairing, while the hairpin handle sequence recruits the Cas9 enzyme. The sequence in the genome targeted by the designed sgRNA needs to be immediately followed by a so-called protospacer adjacent motif (PAM). This conserved 2- to 4-nucleotide sequence is specific for and differs between the various Cas proteins. The PAM of S. pyogenes Cas9 is 5’-NGG-3’.

Multiple technologies, including gene expression knockdown, also called CRISPR interference (CRISPRi), have been developed on the basis of the CRISPR/Cas9 system. CRISPRi was established by generating a catalytically inactive variant of the Cas9 endonuclease (dCas9). This mutated enzyme does not generate DSBs but does bind to DNA, as guided by the specific sgRNA used. By sgRNA targeting of dCas9 to specific sites in a gene or its promoter region, access of RNA polymerase to the DNA can be modulated. In this way, transcript elongation or transcript initiation can be blocked, achieving a gene knockdown effect (Figure 6B) [186].
As a very promising multipurpose genome engineering tool, it is no surprise that the CRISPR/Cas9 system was also introduced for LAB. CRISPR/Cas9 has been applied to several LAB genera, such as *Lactobacillus* and *Lactococcus* [183]–[186]. CRISPR/Cas technology has been successfully employed in *L. lactis* using the *S. pyogenes* Cas9 system. It has been used to modify the genome of phages by homologous recombination with engineered DNA repair templates [183]. The CRISPR/Cas9 system can save time but still requires a homologous recombination step using a repair template, as *L. lactis*, like most prokaryotes, lacks a non-homologous end joining (NHEJ) pathway. NHEJ can repair the ds breaks caused by Cas9 in most eukaryotes. Cas9 targeting of chromosomes in some bacteria, such as *E. coli*, has been shown to be lethal when a template for repair through the homology-directed repair (HDR) pathway is not present [187].

In addition to CRISPR/Cas9 mutagenesis, CRISPRi/dCas9 could be a valuable tool for strain engineering in *L. lactis*. It mimics gene absence via gene expression knockdown. Although knocking down of genes cannot replace knockout mutants, CRISPRi/dCas9 has the advantages of a shorter construction time and allowing to study genes that are essential and of which the absence is, thus, lethal. The dCas9 CRISPRi tool was developed and applied in *L. lactis* by using a nisin-inducible promoter (NICE system; see above) [188] to induce both dCas9 and sgRNA genes to repress the target gene *htrA* [189].
Figure 6. Gene engineering by CRISPR/Cas9 and CRISPRi/dCas9 systems. (A) sgRNA leads wild-type Cas9 from \textit{S. pyogenes} to create double strand breaks (DBSs) in the target DNA. The two repair pathways, non-homologous end joining (NHEJ) or homology-directed repair (HDR), generate the desired gene engineering goals. (B) With two mutations in the nuclease active site region, the dCas9 protein is catalytically inactive. Instead of generating DBSs, sgRNA-guided dCas9 can still recognize and bind to its target DNA, blocking RNA polymerase access to the promoter or elsewhere in the gene, disrupting proper gene transcription and leading to a gene knockdown effect.
Thesis outline

In this thesis, we explored the function of the proteins YbeY that (putatively) interact with sRNAs, as well as that of two sRNAs in *Lactococcus lactis*. It starts, in Chapter 1, with a broad overview of the state of the art in the area of research covered in this thesis. In Chapter 2, we designed a number of CRISPR interference (CRISPRi) platforms with high silencing efficiency in *L. lactis*, as was shown by analyzing the effects of silencing of three genes: *acmA*, encoding the major autolysin in *L. lactis*; *ftsZ*, specifying a tubulin FtsZ protein; and *pbp2b*, coding the penicillin-binding protein Pbp2b in *L. lactis*. The chromosome-based CRISPRi platform was ultimately used to investigate the *L. lactis* putative RNA binding protein, YbeY, as the *ybeY* gene, as we show here, is essential. We uncovered the transcriptomic effects of silencing of *ybeY*, and show as a major effect the downregulation of most of the genes of the 30S and 50S ribosomal proteins. In addition, 10 sRNA genes were affected upon the silencing of *ybeY*. In Chapter 3, we studied ZnsQ, an sRNA derived from 3’-untranslated region (3’-UTR) the *zitRSQP* transcript. The *zit* operon is involved in Zn(II) uptake in *L. lactis*. Combining *in silico* target prediction and MS2 affinity purification coupled with RNA sequencing (MAPS) technology, we explored the mRNA targets of ZnsQ and show that it is involved in Zn(II) uptake. It targets the transcript of the Zn(II) transporter genes *zitQP* and the *zit* operon repressor gene *zitR* to lower the Zn(II) uptake capacity in order to prevent zinc toxicity. It also provides a quick-start once Zn(II) becomes limiting. In Chapter 4, S092, an sRNA derived from the intergenic region between the genes *thiE* and *dltA* of *L. lactis* MG1363, was investigated by following the growth of cells under a number of stress conditions, on various sugars as the sole carbon source, and combining that with light microscopy observations. Also, RNA-seq was performed on a strain overexpressing S092. Using MAPS, direct target genes of S092 were pinpointed. Our findings reveal that S092 may be involved in carbon metabolism and cell envelope stress response. Chapter 5 presents a summary and discussion of the work presented in this study, putting forward perspectives for future research in this area. We also present a primary exploration of *L. lactis* RapZ, as it was recently proposed to be able to bind sRNAs. RNA-seq analysis of *rapZ* overexpression and deletion strains suggest that RapZ takes part in sugar metabolism. It might also be involved in *arc* operon regulation by stabilizing ArgX or stimulating its function. Via co-immunoprecipitation (Co-IP) analysis, we identified several proteins that can be pulled down together with RapZ.
REFERENCES


Utilization in Lactococcus lactis, ”J. BACTERIOLOGY, 176, 5681–5685.


Chapter 1


derived from the 3′-UTR region of its gene, ArgX, both regulate the arginine deiminase pathway in *Lactococcus lactis*, "*PLoS One*, 14, 512-518.


