



universität  
wien

# DIPLOMARBEIT

Titel der Diplomarbeit

## **Analysis of novel small non-coding RNAs in *Streptococcus pyogenes***

angestrebter akademischer Grad

Magister/Magistra der Naturwissenschaften (Mag. rer.nat.)

Verfasserin / Verfasser:	Doris Veit
Matrikel-Nummer:	0203387
Studienrichtung (lt. Studienblatt):	Molekulare Biologie
Betreuerin / Betreuer:	Dr. Emmanuelle Charpentier

Wien, am 13.08.2008

# Table of contents

---

Table of contents .....	2
1. Abstract .....	5
2. Zusammenfassung .....	7
3. Introduction .....	9
3.1 <i>Streptococcus pyogenes</i> .....	9
3.1.1 <i>S. pyogenes</i> virulence factors .....	10
3.1.1.1 Cell-associated virulence factors .....	10
3.1.1.2 Secreted virulence factors .....	11
3.1.2 Regulation of virulence factors .....	12
3.1.2.1 Stand alone response regulators .....	13
3.1.2.1.1 Multiple gene regulator (Mga) .....	13
3.1.2.1.2 RofA-like protein (RALP) .....	13
3.1.2.1.3 Rgg/RopB (Regulation of proteinase) .....	14
3.1.2.2 Two component systems (TCS) .....	14
3.1.2.2.1 CovR/S (control of virulence genes) .....	14
3.1.2.2.2 FasBCAX (fibronectin/fibrinogen binding/hemolytic activity/streptokinase regulator) .....	15
3.1.2.2.3 Ihk-Irr (isp-adjacent histidine kinase/response regulator) .....	15
3.1.2.3 Regulatory sRNAs in <i>S. pyogenes</i> .....	15
3.1.2.3.1 <i>pel</i> RNA .....	16
3.1.2.3.2 <i>fasX</i> RNA .....	16
3.1.2.3.3 <i>RivX</i> RNA .....	16
3.2 Regulatory RNAs in bacteria .....	17
3.2.1 Mode of action of sRNAs .....	19
3.2.1.1 Base-pairing mechanism .....	20
3.2.1.1.1 <i>Cis</i> -encoded sRNAs .....	20
3.2.1.1.2 <i>Trans</i> -encoded sRNAs .....	21
3.2.1.2 RNA-Protein interaction .....	22
3.2.1.3 RNAs with intrinsic enzymatic activity .....	22
3.2.2 Some additional examples of sRNAs involved in the control of bacterial pathogenicity .....	23
3.2.2.1 RNAIII in <i>Staphylococcus aureus</i> .....	23
3.2.2.2 Qrr1-Qrr4 in <i>Vibrio cholerae</i> .....	23
3.2.2.3 RNA $\alpha$ in <i>Vibrio anguillarum</i> .....	24
3.2.3 Regulatory RNA elements .....	24
3.2.3.1 Riboswitches .....	24
3.2.3.2 T boxes .....	27
3.2.3.3 Leader RNAs .....	27
3.2.3.4 Housekeeping RNAs .....	28
3.2.3.4.1 6S RNA .....	28
3.2.3.4.2 tmRNA .....	28
3.2.3.4.3 RNase P .....	28

3.2.3.4.4 SRP .....	29
<b>4. Materials and Methods .....</b>	<b>30</b>
4.1 Computational predictions and analysis of sRNAs.....	30
4.1.1 RNA predictions using computational tools .....	30
4.1.2 Preparation of probes for transcriptional expression analysis.....	31
4.1.3 Secondary structure predictions.....	31
4.1.4 mRNA target prediction .....	31
4.2 Bacterial strains, growth conditions and media.....	31
4.2.1 Bacterial strains .....	31
4.2.2 Growth conditions and media .....	32
4.2.2.1 <i>S. pyogenes</i> .....	32
4.2.2.2 <i>Escherichia coli</i> .....	32
4.2.2.3 Antibiotics.....	33
4.3. DNA preparation.....	33
4.3.1 Plasmid preparation from <i>E. coli</i> .....	33
4.3.1.1 Quick gram-negative plasmid preparation.....	33
4.4 Standard DNA manipulation procedures.....	34
4.4.1 Polymerase chain reaction (PCR) .....	34
4.4.1.1 Taq-DNA polymerase .....	34
4.4.1.2 Phusion polymerase .....	34
4.4.2 Restriction digestions .....	34
4.4.3 Purification of DNA fragments.....	35
4.4.4 Agarose gel electrophoresis .....	35
4.4.5 General cloning steps .....	35
4.4.6 Preparation of competent <i>E. coli</i> cells.....	36
4.4.7 Heat-shock transformation of competent <i>E. coli</i> cells.....	36
4.4.8 Measurement of DNA concentration.....	36
4.5 Primer extension .....	37
4.5.1 PCR product sequencing .....	37
4.5.2 $\gamma$ - <sup>32</sup> ATP Labelling of primers .....	37
4.5.3 Primer extension .....	37
4.5.3.1 Primer extension using Promega AMV Reverse Transcriptase .....	37
4.5.3.2 Primer extension using Invitrogen ThermoScript Reverse Transcriptase .....	38
4.5.4 Sequencing gel.....	38
4.6 RNA techniques .....	39
4.6.1 Preparation of total RNA from <i>S. pyogenes</i> .....	39
4.6.2 PAA gel electrophoresis.....	40
4.6.3 Northern blot analysis using PAA gels .....	41
4.6.4 Northern Blot using agarose gels .....	42
4.6.5 Labelling (Random Priming) of probes .....	43
4.6.6 Determination of RNA stability.....	43
4.7 StreptoTag.....	44
4.7.1 <i>In vitro</i> transcription of tagged RNA .....	44
4.7.2 Total cell lysates .....	44

4.7.3 SDS-PAGE analysis.....	45
<b>5. Aims of the study .....</b>	<b>47</b>
<b>6. Result.....</b>	<b>49</b>
6.1 Selection of sRNA candidates for experimental validation and analysis (preliminary data).....	49
6.2 Experimental validation of expression of sRNA candidates using a Northern blot screen.....	49
6.3 Mapping of 5' and 3' extremities of sRNA.....	55
6.4 Stability of sRNAs.....	58
6.5 Computational analysis of sRNAs .....	60
6.5.1 Structure predictions .....	60
6.5.2 sRNA target predictions .....	61
6.6 Further analysis of novel sRNA candidates.....	63
6.6.1 Comparative expression analysis of sRNA candidates with mRNAs of adjacent ORFs: the case of SpyRNA014.....	63
6.6.2 Genomic localization analysis of novel sRNA candidates.....	65
6.6.3 The case of SpyRNA049 .....	65
6.6.3.1 Genomic localization and expression analysis of SpyRNA049 .....	65
6.6.3.2 Creation of a SpyRNA049 mutant.....	66
6.6.3.4 Fishing proteins binding SpyRNA049 using the StreptoTag method....	68
<b>7. Discussion .....</b>	<b>71</b>
7.1 Computational approach to search for novel sRNAs in <i>S. pyogenes</i> and experimental validation using Northern blot screen.....	71
7.2 Characterization of expressed sRNA candidates .....	72
7.3 Analysis of SpyRNA049.....	74
7.4 Additional analysis of SpyRNA014.....	76
7.5 Outlook.....	76
<b>8. Appendix A .....</b>	<b>78</b>
<b>9. Appendix B.....</b>	<b>96</b>
<b>10. References.....</b>	<b>127</b>
<b>11. Acknowledgements .....</b>	<b>135</b>
<b>12. Curriculum Vitae .....</b>	<b>136</b>
<b>13. Lebenslauf .....</b>	<b>137</b>

# 1. Abstract

---

*Streptococcus pyogenes* (*S. pyogenes*) is an important human pathogen that causes a wide range of diseases varying from mild infections like pharyngitis, impetigo or scarlet fever to severe life threatening invasive diseases like the streptococcal toxic shock syndrome (STSS) or necrotising fasciitis. To establish and maintain an infection, *S. pyogenes* utilizes a large panel of virulence factors that are either secreted or associated with the cell surface of the bacterium. In order to regulate the expression of these virulence factors this versatile pathogen uses protein regulators such as two-component systems and stand alone response regulators as well as newly described small regulatory RNAs (sRNAs). So far, three of such sRNAs have been identified in *S. pyogenes*: *fasX*, *pel* and *rivX* RNA.

The aim of this study was to identify novel small non-coding RNAs in *S. pyogenes*. Using a large-scale computational approach, intergenic regions (IGRs) likely to encode sRNAs in the *S. pyogenes* genome were identified and summarized into 178 loci. Here, 90 (50.6%) were further studied using Northern blot analysis. Under normal growth conditions 29 (~32%) were expressed. Five sRNAs were predicted riboswitches, six were leader elements, four were functional RNAs, seven were predicted T boxes, five were novel sRNAs of unknown function and SpyRNA014 was already once mentioned in the literature. The 5' and 3' extremities of the sRNAs were mapped and their stability was tested in rifampicin experiments. *In silico* predictions of the secondary structure of some of the sRNA candidates were done. A computational program was also used to identify possible messenger RNA (mRNA) targets of six sRNA candidates. The sRNA candidate SpyRNA049 was chosen for further analysis of its possible function. This included (i) starting the construction of a SpyRNA049-deficient strain in an M1 *S. pyogenes* strain and (ii) applying the 'StreptoTag' biochemical method to fish putative protein targets binding to SpyRNA049.

Taken together, novel sRNA have been identified in *S. pyogenes*, most of them being part of 5' untranslated regions of genes. Because of their location in the

genome and their homologies with already described sRNAs, these sRNAs are likely to play crucial roles in metabolic and pathogenicity-linked processes in *S. pyogenes*.

## 2. Zusammenfassung

---

*Streptococcus pyogenes* (*S. pyogenes*) ist einer der wichtigsten menschlichen Krankheitserreger, der ein breites Spektrum an Krankheiten verursacht, beginnend bei leichten Infektionen wie Hals- und Rachenentzündungen, Impetigo oder Scharlach, bis zu schweren invasiven Krankheiten wie das Toxische Schock Syndrom oder nekrotisierende Fasciitis, die auch lebensbedrohlich sein können. Um eine Infektion zu etablieren und dann auch aufrecht zu erhalten, besitzt *S. pyogenes* eine große Anzahl von Virulenzfaktoren, die sich entweder auf der Zelloberfläche befinden oder vom Bakterium in die Umgebung freigesetzt werden. Die Expression dieser Virulenzfaktoren wird einerseits durch protein-kodierte Zwei-Komponentensysteme und Transkriptionsfaktoren, aber auch durch erst kürzlich beschriebene RNA Moleküle reguliert. Von diesen RNAs, auch „small RNAs“ (sRNAs) genannt, sind in *S. pyogenes* bis jetzt drei bekannt, nämlich *fasX*, *rivX* und *pel* RNA.

Das Ziel dieser Studie war die Identifizierung neuer nicht-kodierender sRNAs in *S. pyogenes*. Dafür wurden mittels Computeranalyse Regionen zwischen einzelnen Genen ermittelt, die für potenzielle sRNAs kodieren. Von den insgesamt 178 gefundenen Loci wurden 90 mittels Northern blot Analyse weiter untersucht. Dabei konnte die Expression von 29 sRNAs (~32%) unter normalen Wachstumsbedingungen nachgewiesen werden. Diese 29 Kandidaten konnten in 5 funktionelle Kategorien eingeteilt werden: (i) riboswitch (5), (ii) leader elements (6), (iii) funktionelle RNAs (4), (iv) T boxes (8), (v) sRNAs mit unbekannter Funktion (5) und (vi) SpyRNA014, welche bereits einmal in der Literatur als sRNA erwähnt wurde. Des Weiteren wurden die 5' und 3' Enden, sowie die Stabilität der sRNAs mittels Rifampicin Experimenten bestimmt. Für einige der Kandidaten wurde mittels Computeranalyse eine Vorhersage über mögliche Sekundärstrukturen erstellt und zusätzlich konnten bei sechs Kandidaten mögliche messenger RNAs (mRNAs) als Bindungspartner vorhergesagt werden.

Kandidat SpyRNA049 wurde für die weitere funktionelle Analyses ausgewählt. Diese beinhaltete (i) den Beginn der Konstruktion einer Mutante, sowie (ii) die Anwendung der biochemischen Methode „StreptoTag“ um mögliche Proteine, die an SpyRNA049 binden, zu fischen.

Während dieser Arbeit konnten neue sRNAs in *S. pyogenes* identifiziert werden, von denen die meisten in 5' untranslatierten Regionen liegen. Aufgrund ihrer Lage im Genom und der Homologie mit bereits beschriebenen sRNAs, ist es wahrscheinlich, dass diese sRNAs eine wichtige Rolle in metabolischen oder virulenzbezogenen Prozessen spielen.

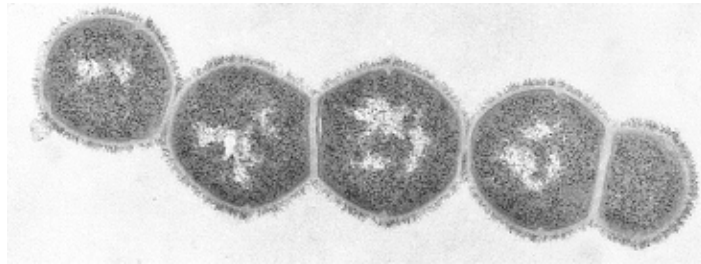


## 3. Introduction

---

### 3.1 *Streptococcus pyogenes*

*Streptococcus pyogenes* (*S. pyogenes*) or 'Group A streptococcus' (GAS) (Fig. 1) is a gram-positive human pathogen that belongs to the family of homofermentative lactic acid bacteria. It is facultatively anaerobe, catalase negative and has  $\beta$ -hemolytic activity. *S. pyogenes* is non-motile and grows in pairs or chains (Baron and National Center for Biotechnology Information (U.S.), 1996). In 1928, Dr. Rebecca Lancefield described a system by which different serotypes (M types) of *S. pyogenes* can be distinguished using antibodies specific to a hypervariable region of a cell-surface protein called the 'M-protein' (Facklam *et al.*, 2002).



**Fig. 1.** Electron micrograph of group A streptococci growing in a chain.  
(from <http://www.rockefeller.edu/vaf/chain20.htm>, last accessed 15.05.2008)

GAS is one of the most common pathogens, it can only infect humans and causes a wide range of diseases. GAS infections vary in their severeness from mild diseases like pharyngitis (strep throat), impetigo and scarlet fever to severe invasive infections such as myositis, necrotising fasciitis (flesh eating disease) and streptococcal toxic shock syndrome (STSS), which even nowadays have high mortality rates. It has been suggested that certain serotypes are associated with skin infections, like M type 49 (M49) and others with throat infections, such as M1 (Bisno, 1995). Furthermore, post infection immune sequelae can develop, such as glomerulonephritis, arthritis, acute rheumatic fever and rheumatic heart

disease, the latter two diseases being the most serious autoimmune sequelae (Cunningham, 2008).

The genome of the M1 reference strain SF370, which was the first to be sequenced (Ferretti *et al.*, 2001) and used throughout this study, is 1,852,441 bp long. This genome encodes 1811 genes that have been annotated (reaching 83% of the genome) and 17% of intergenic regions (NCBI, 2008).

### **3.1.1 *S. pyogenes* virulence factors**

*S. pyogenes* is a versatile bacterium, which utilizes a number of virulence factors necessary for the establishment or maintenance of an infection. These factors promote colonization, allow evasion of the immune system and permit survival in the harsh host environment. Pathogenicity of *S. pyogenes* is multifactorial and no association of a specific virulence factor with a specific disease has been described so far. The virulence factors in *S. pyogenes* are divided into two groups: (i) cell-associated and (ii) secreted virulence factors (Bisno *et al.*, 2003).

#### **3.1.1.1 Cell-associated virulence factors**

The first step for successful infection is the adherence of the pathogen to the host surface, such as skin or throat surface, which are the most important primary infection sites of *S. pyogenes*. The best-known adherence molecules of *S. pyogenes* are lipoteichoic acid, fibronectin binding proteins and the M protein. The M protein, encoded by the *emm* gene, is anchored in the cell wall of *S. pyogenes* (Fischetti, 1991). It consists of two polypeptide chains that form a  $\alpha$ -helical coiled-coil structure. The Lancefield serological classification of GAS is based on detecting differences in the hypervariable region of the M protein among the different GAS strains. The M protein prevents the opsonisation via the alternate complement pathway and helps the bacterium to escape phagocytosis by cells of the human immune system, but also triggers the uptake of the bacterium by non-phagocytic host cells. Apart from the M protein and M-like proteins, the bacterial capsule also supports the prevention of internalisation by host immune cells. The

capsule of *S. pyogenes* consists of hyaluronic acid, which is chemically similar to human connective tissue and therefore is not highly immunogenic. The encapsulation is regulated by the *has* operon, which consists of three genes that are important for the synthesis of hyaluronic acid: *hasA* encoding a hyaluronate synthase, *hasB* encoding an UDP-glucose dehydrogenase and *hasC*, encoding an UDP-glucose pyrophosphorylase. Different *S. pyogenes* strains are encapsulated to different extents. A higher degree of encapsulation has been implicated with increased virulence, because of an enhanced ability to resist phagocytosis by cells of the human immune system (Bisno *et al.*, 2003).

Fibronectin binding proteins are important for the adherence to skin and throat and include protein F1 (PrfF1/SfbI), SfbII, FBP54, PFBP and protein F2 (Bisno *et al.*, 2003).

The C5a-peptidase is another cell-surface associated enzyme that cleaves the C5a component of the complement system, preventing its chemoattracting effect on polymorphonuclear leukocytes (PMN) and macrophages (McIver and Scott, 1997).

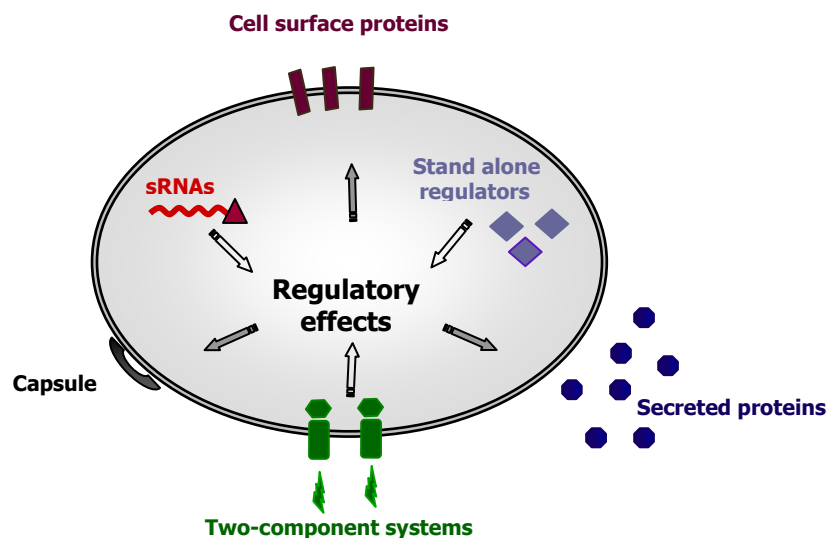
### **3.1.1.2 Secreted virulence factors**

*S. pyogenes* also releases several extracellular virulence factors. It produces two pore forming cytolytins: Streptolysin O (SLO), which is reversibly inhibited by oxygen and Streptolysin S (SLS), which is not sensitive to oxygen but thermolabile. SLS exists in three forms: intracellular, extracellular and membrane bound. Both SLO and SLS can damage membranes of certain blood cells and subcellular organelles. SLS is responsible for the typical  $\beta$ -haemolysis that can be observed on blood agar plates. Additional secreted virulence factors are DNases, hyaluronidase, streptokinase, the cysteine protease SpeB and the streptococcal inhibitor of complement (Bisno *et al.*, 2003; Hynes, 2004). Other potent virulence factors are the streptococcal pyrogenic exotoxins, which are bacterial superantigens that seem to play an important role in STSS. These exotoxins stimulate an unspecific T cell proliferation to a great extent. Superantigens have

the ability to bind simultaneously to T cell receptors and MHC II molecules, thus stimulating proliferation of T cells and an extensive release of cytokines, like TNF- $\alpha$ , through the T cells. This strong inflammatory reaction leads to a reduced vascular tone, which can result in decreased blood supply of organs, acidosis or multiple organ failure (Cunningham, 2000; Sriskandan *et al.*, 2007).

### 3.1.2 Regulation of virulence factors

During an infection, pathogens have to adapt to a wide variety of different conditions within the host, among them nutrient limitation or evasion from the host immune system. In these processes, expression of virulence factors need to be tightly controlled. Regulation of virulence gene expression occurs via protein regulators such as transcriptional regulators or two component systems, but also via RNA regulators (Fig. 2).



**Fig. 2.** Simplified overview of virulence factor regulation in *S. pyogenes*. Three main regulatory elements, two-component systems, 'stand-alone' regulators and sRNAs, control the expression of a large set of virulence factors consisting of cell surface proteins, secreted proteins and the capsule.

### 3.1.2.1 Stand alone response regulators

For all 'stand alone' response regulators, it is still unclear how they exactly sense environmental signals (Kreikemeyer *et al.*, 2003). No interacting sensing molecules have been identified yet.

#### 3.1.2.1.1 Multiple gene regulator (Mga)

Multiple gene regulator (Mga) is a typical 'stand-alone' response regulator acting as DNA-binding protein. Mga regulates the expression of several virulence factors, such as the M protein (*emm*), M-like Ig proteins (*mrp*, *arp* and *enn*), serum opacity factor (*sof*), C5a peptidase (*scpA*) and the secreted inhibitor of complement (*sic*), by binding to their upstream promoter region. Mga also regulates its own expression (Hondorp and McIver, 2007). The 62 kDa protein is strongly expressed during exponential growth and seems to indirectly regulate the expression of other operons that are important for metabolism, sugar transport and sugar utilisation (Ribardo and McIver, 2006). The Mga regulon (*vir* regulon) is present in all *S. pyogenes* strains, although its genetic arrangement is not similar in all serotypes. Expression of Mga seems to be regulated according to environmental signals (temperature, CO<sub>2</sub> and iron levels) and it is repressed upon entry into stationary phase by RALPs and Rgg/RopB (Hondorp and McIver, 2007).

#### 3.1.2.1.2 RofA-like protein (RALP)

The M6 serotype of GAS expresses RofA, a DNA-binding protein that positively regulates the *prtF1* gene encoding the protein F1, probably the most important fibronectin-binding molecule expressed (Fogg *et al.*, 1994; Fogg and Caparon, 1997). In other M types homologous proteins to RofA were identified leading to the establishment of the term RofA-like protein (RALP) family (Beckert *et al.*, 2001). These proteins ensure a reduced virulence factor expression for intracellular long-term survival and stationary phase of growth. RALPs regulate the expression of a set of genes that include MSCRAMM (microbial surface

components that recognize adhesive matrix molecules, like fibronectin or collagen binding proteins) genes, *sagA* (SLS), *speB* (cysteine protease), *speA* (superantigen) and *mga* (Kreikemeyer *et al.*, 2003).

#### **3.1.2.1.3 Rgg/RopB (Regulation of proteinase)**

Rgg/RopB is a DNA-binding transcriptional regulator, which influences the expression of *SpeB*, *Mga* and the two component systems *CsrRS/CovRS*, *FasBCAX* and *Ihk/Irr*. Since Rgg/RopB is negatively regulating *Mga* transcription, it might play a role in reducing or even preventing the *Mga* expression during stationary phase. Rgg/RopB regulators are conserved among gram-positive bacteria (Chaussee *et al.*, 2002).

#### **3.1.2.2 Two component systems (TCS)**

TCS are common systems in bacteria which allow the sensing of extracellular signals and mediate adaptation of gene expression according to environmental changes. In GAS, 13 TCS were identified upon sequencing of the genome (Ferretti *et al.*, 2001). Common features of TCS are that they consist of at least one sensing molecule, which is a transmembrane histidine kinase, and an intracellular response regulator. In general, binding of an extracellular signal activates the sensor histidine kinase. This leads to dimerization of the receptors upon activation and autophosphorylation in the transducer domain. The activated transducer domain catalyses the phosphorylation of the response regulator that directly regulates gene transcription (Kreikemeyer *et al.*, 2003).

##### **3.1.2.2.1 CovR/S (control of virulence genes)**

*CovR/S*, also referred to as the *CsrR/S* (capsule synthesis regulator) system is the best described TCS in GAS and has been shown to directly or indirectly influence about 15% of the *S. pyogenes* genome. This includes several virulence factors like the *has* operon, *ska*, *speB*, *sagA* and *RivR* (RALP4). In this system, *CovR* is the response regulator and it predominantly represses transcription of the regulated

genes. The CovRS system is mainly active during late exponential growth and stationary survival (Churchward, 2007). The CovRS systems seems to regulate virulence factor expression according to  $Mg^{2+}$  levels. When the  $Mg^{2+}$  concentration is low, the covRS system represses the expression of virulence factors, while at high  $Mg^{2+}$  concentrations that would correspond to intracellular conditions of the human body it switches to an activating role (Gryllos *et al.*, 2003).

#### **3.1.2.2.2 FasBCAX (fibronectin/fibrinogen binding/hemolytic activity/streptokinase regulator)**

A second prominent TCS is the FasBCAX system that consists of two putative sensor histidine kinases (FasB and FasC), one putative response regulator (FasA) and as effector molecule of the system, the sRNA *fasX*. The FasBCAX operon is expressed in a growth phase dependent manner. It downregulates the expression of fibrinogen and fibronectin binding proteins and it has a positive regulatory effect on the expression of secreted virulence factors like streptokinase or streptolysin S (SLS) (Kreikemeyer *et al.*, 2001; Kreikemeyer *et al.*, 2003; Klenk *et al.*, 2005). Please refer to section 3.1.2.2.2 for details on *fasX* RNA.

#### **3.1.2.2.3 Ihk-Irr (isp-adjacent histidine kinase/response regulator)**

The Ihk-Irr TCS has an important role in the escape of GAS from polymorphonuclear leukocyte (PMN) killing and assists host-cell lysis. It has been shown that there is a high expression of the Ihk-Irr TCS in GAS causing acute pharyngitis (Voyich *et al.*, 2003).

#### **3.1.2.3 Regulatory sRNAs in *S. pyogenes***

Please refer to section 3.2 for details on small regulatory RNAs (sRNAs).

#### 3.1.2.3.1 *pel* RNA

The pleiotropic effect locus (*pel*) region contains the *sagA* gene that encodes the haemolytic enzyme SLS and an untranslated RNA that regulates the expression of virulence factors. *pel* RNA regulates the expression of *emm* (M-protein), *sic* (streptococcal inhibitor of complement) and *nga* (N-acetylglucosamine) at the transcriptional level and acts posttranscriptionally on SpeB. *pel* RNA (459 nt long) is transcribed in a growth phase-dependent manner and its transcription is induced by conditioned media, which indicates a possible regulation of *pel* RNA by other components. RNAlII and *pel* RNA share several characteristics: they are both encoding a peptide, while the RNA itself has a regulatory function, they have approximately the same size and similar thermodynamic structure predictions. The mechanisms of *pel* RNA effects are so far unknown (Mangold *et al.*, 2004).

#### 3.1.2.3.2 *fasX* RNA

As mentioned earlier, *fasX* is the effector molecule of the *fasBCA* (fibronectin/fibrinogen binding/haemolytic activity/streptokinase regulator) operon. Conserved among different M types *fasX* is transcribed into a 200 nt long RNA molecule under normal growth conditions, while under amino acid starvation *fasX* is transcribed from a promoter lying further upstream into a 300 nt long RNA (Steiner and Malke, 2001). *fasX* is predominantly expressed during mid- to late-log growth under normal growth conditions (Kreikemeyer *et al.*, 2001). A contribution of *fasX* RNA to GAS virulence has been shown with a *fasX* RNA deficient strain that revealed that this mutant strain is less virulent in terms of adherence and internalization to epithelial cells (Klenk *et al.*, 2005).

#### 3.1.2.3.3 *RivX* RNA

*RivX* is an sRNA encoded downstream of the *rivR* gene, which encodes a regulator of the RALP family. The operon that encodes *rivX* RNA and RivR is a target of the CovR repressor. The two genes, *rivR* and *rivX*, are cotranscribed and



seem to regulate independently genes of the *mga* regulon that are involved in virulence such as *emm*, *scpA*, *speB* or *mga*. *rivX* RNA was shown to act as a non-coding RNA. Mutating the putative start codons within the *rivX* sequence did not change its regulatory properties. *rivX* RNA has three different 5' ends, which were suggested to be produced through processing of the *rivRX* transcript. RivR and *rivX* RNA are the first regulators found that provide a link between Mga and the CovRS system (Roberts and Scott, 2007).

### 3.2 Regulatory RNAs in bacteria

During the past few years, numerous RNAs other than messenger RNAs (mRNAs), ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) have been reported in bacteria, as well as in other kingdoms of life. These RNAs possess regulatory properties and several names have been used in the literature to define them: 'small RNAs' (sRNAs) mainly used in bacteria (term chosen in this study) or 'non-coding RNAs' (ncRNAs) mainly used for eukaryotic sRNAs. The term 'regulatory RNAs' (Storz and Haas, 2007) is also employed since exceptions of RNAs (e.g. the *Staphylococcus aureus* RNAIII) that are long and encode a peptide were reported, thus contradicting the terms sRNA or ncRNA (Toledo-Arana *et al.*, 2007). In general, regulatory RNAs seem to enable bacteria to fine-tune their adaptive response to varying environmental conditions (Altuvia and Wagner, 2000).

Several methods have been described, which were used to identify unknown sRNAs in bacteria, although many of the sRNAs were identified by chance. For review please refer to Hüttenhofer and Vogel (2006) and Altuvia (2007). One of these systematic approaches is the direct sequencing of sRNAs. For this procedure the RNAs need to be very abundant in a cell. The method relies on running bacterial total RNA on a polyacrylamide (PAA) gel and excision of a sRNA single band from the ethidium bromide stained gel, which is subsequently used for sequencing procedure. To identify less abundant sRNAs the RNAs can be labelled before loading on the gel by metabolic labelling, or by 3' or 5'

labelling before sequencing. Another strategy, named RNomics, is to produce a specialized cDNA library, from sRNAs that are first size selected using polyacrylamide gel electrophoresis (PAGE), reverse transcribed to cDNA (using linker and adaptor oligonucleotides) and then cloned into standard vectors for further sequencing or directly sequenced using high throughput sequencing technology (454 or selexa) (Kawano *et al.*, 2005; Vogel *et al.*, 2003). Genomic SELEX is an additional method based on the property of sRNAs to interact specifically with proteins. This method is suitable for bacterial species containing the RNA chaperone Hfq, which is known to bind sRNAs and can therefore be used as protein trap. However there are some bacterial species in which no specific sRNA-binding protein has been identified yet. This is the case for *S. pyogenes*, which does not seem to encode an Hfq homologue. For these bacterial species, this method is not suitable. The last method to mention are microarrays, which were specifically designed to analyse sRNA expression, because so far commercially available microarrays did only contain probes for ORFs. Tiling microarrays carrying overlapping oligonucleotide probes for both strands of intergenic regions (IGRs) in addition to ORFs were for example used in the study of Wassarman *et al.* (2001). Of note are the recently available antibodies that specifically detect DNA-RNA hybrid molecules and overcome the problem of labelling nucleic acids (Hüttenhofer and Vogel, 2006). The most common experimental methods that are currently used to identify sRNAs are RNomics and tiling-arrays to identify novel sRNAs.

Beside experimental approaches, computational tools were used to identify sRNAs in several bacterial pathogens (Argaman *et al.*, 2001; Rivas and Eddy, 2001; Rivas *et al.*, 2001; Lenz *et al.*, 2004; Wilderman *et al.*, 2004; Livny *et al.*, 2006; Gonzalez *et al.*, 2008). The typical features that are taken in account for computational approaches are conservation in structure or sequence as well as transcription regulation sequences and other typical genomic features.

### 3.2.1 Mode of action of sRNAs

Many different regulatory functions have been reported for sRNAs (Gottesman, 2004) affecting different levels of gene expression. The modes of action of the sRNAs can be divided into three main mechanisms: (i) base-pairing with target nucleic acid molecules, (ii) RNA-protein interactions (Fig. 3) or (iii) enzymatic activity of RNAs (Storz *et al.*, 2004).

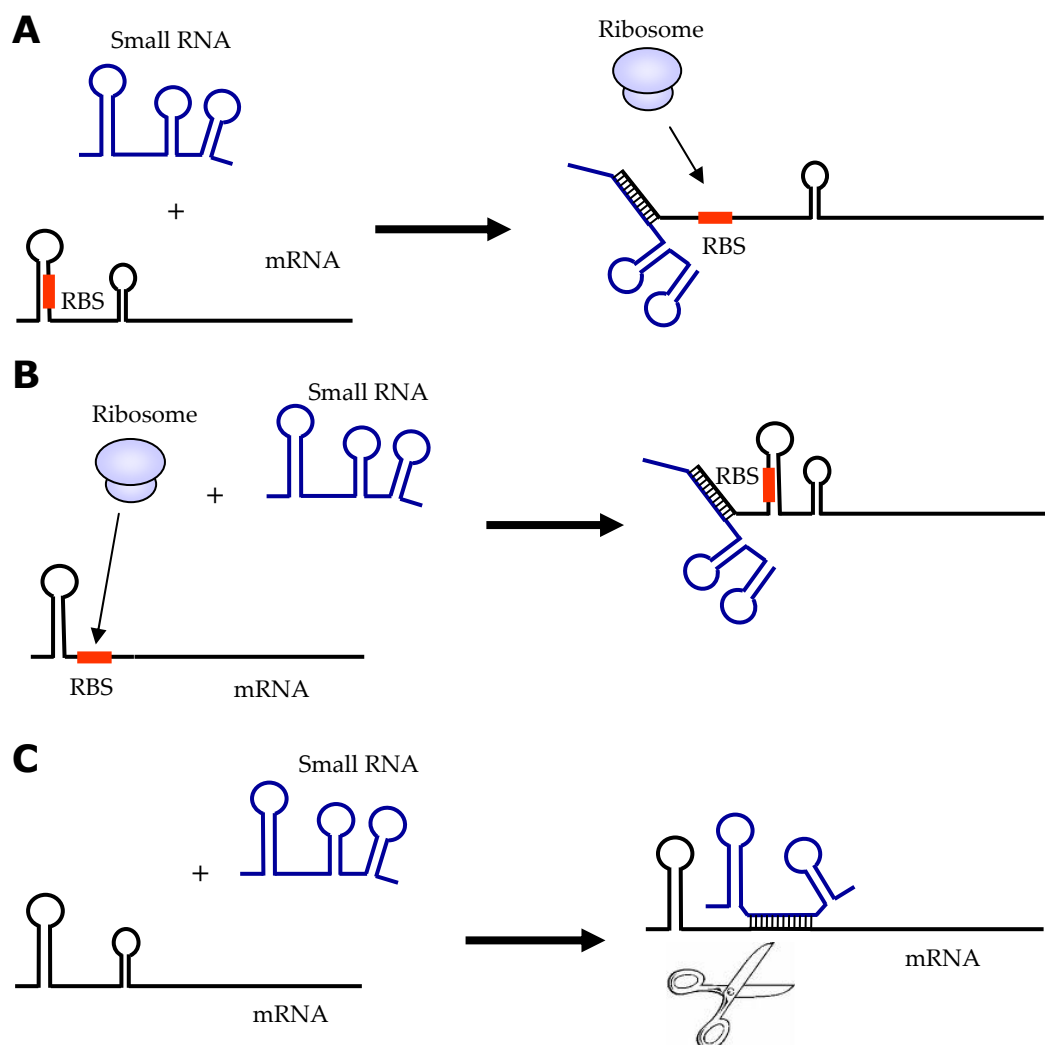


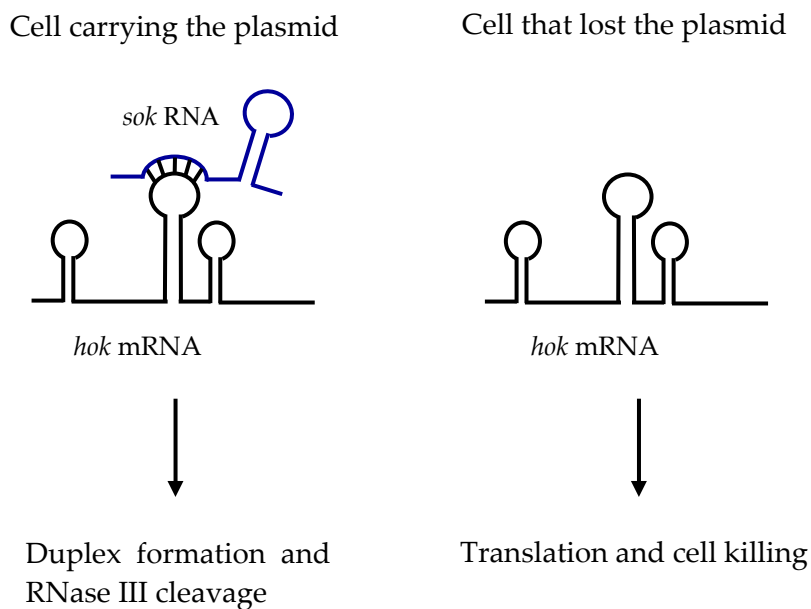
Fig. 3. Examples of sRNAs (blue) basepairing mechanisms with target mRNAs (black). Upon RNA-RNA interaction the ribosomal binding site (RBS) (red) can be released and ribosome binding is enabled (A) or the RBS is sequestered and translation is inhibited (B). Basepairing of sRNAs with mRNAs can also lead to degradation of the complex (C). Figure taken from (Storz *et al.*, 2004).

### 3.2.1.1 Base-pairing mechanism

The most frequently reported mode of action for sRNAs is the base-pairing mechanism (Storz *et al.*, 2005). These antisense regulators can be encoded in *cis* or in *trans* to their mRNA targets.

#### 3.2.1.1.1 *Cis*-encoded sRNAs

*Cis* acting sRNAs are encoded in the same genomic locus and are therefore usually perfectly complementary to their target, being encoded on the opposite strand. They were first described in studies focussing on plasmid replication control in toxin-antitoxin systems like the *hok/sok* system (Fig.4). In this system, the *hok* gene encodes a toxin that causes cell membrane damage and the *sok* RNA is an unstable sRNA that inhibits indirectly *hok* mRNA translation. The *hok* mRNA is stable and gets translated only upon loss of the plasmid R1, encoding the *hok/sok* locus. In this case, the inhibitory effect of *sok* RNA is abolished and the stable *hok* mRNA persists in the cell (Gerdes and Wagner, 2007).



**Fig. 4.** *hok/sok* system. Figure modified from (Gerdes and Wagner, 2007).

### 3.2.1.1.2 *Trans*-encoded sRNAs

sRNAs encoded in *trans* to their target mRNAs do not show perfect complementarity and usually target more than one mRNA. These sRNAs control the stability and/or the translation of the target mRNAs.

In *E. coli* many sRNAs (OxyS, Spot42, DsrA, RhyB...) that are encoded in *trans*, need the RNA chaperone Hfq (host factor 1) for their function (Brennan and Link, 2007). This protein assists the interaction between mRNA and sRNA. A common mechanism of *trans*-encoded sRNAs is the regulation of translation of the target mRNAs by affecting the access of the ribosome to the ribosome binding site (RBS) (Fig. 4A and 4B). As best example, OxyS RNA is a sRNA in *E. coli*, 109 nt long and strongly expressed in cells submitted to oxidative stress. Using a base-pairing mechanism, OxyS RNA represses the translation of the *fhfA* (encoding a transcriptional activator) and *rpoS* (encoding the stationary phase  $\sigma$ -factor) by sequestering the RBS (Gottesman, 2004; Liu *et al.*, 2005). A second example is DsrA, an sRNA that is important for gene regulation at lower temperatures. DsrA RNA seems to allow an increased translation of *rpoS* mRNA via a base-pairing mechanism (Majdalani *et al.*, 2002). As third example, the sRNA Spot42 is involved in the regulation of sugar metabolism in *E. coli*, by promoting discoordinate expression of the genes within the *gal* operon at the translational level (Moller *et al.*, 2002).

The RNA-RNA interaction in sRNA-mRNA complexes can also influence the stability of mRNAs, where the complex can be the direct target for degradation by RNases (Fig. 4C). As example this mechanism was suggested for RhyB RNA in *E. coli*, where an increased level of RNase E-mediated digestion was observed upon RhyB basepairing with its target mRNAs (Masse *et al.*, 2003). Via this mechanism, RhyB down-regulates the expression of many mRNAs encoding iron containing proteins (Masse and Gottesman, 2002). On the contrary it was also suggested that base pairing might stabilise sRNA-mRNA complexes and prevent degradation (Gottesman, 2004; Storz *et al.*, 2005).

### 3.2.1.2 RNA-Protein interaction

The second main mechanism of regulation is an interaction of the sRNA with proteins. Upon binding of the sRNA to the protein, the activity of the protein can be altered, or proteins can be quickly sequestered, which results in the blockage of their function. One example of a sRNA binding to a protein is the 6S RNA in *E. coli*. This sRNA sequesters  $\sigma^{70}$ -RNA-polymerase by mimicking the structure of open promoter DNA and thereby down-regulates transcription by competing with promoter DNA for  $\sigma^{70}$ -RNA-polymerase binding (Wassarman, 2007). Please refer to paragraph 3.2.3.4.

Examples of this mechanism include also the sRNAs, CsrB and CsrC in *E. coli* and RsmY and RsmZ in *Pseudomonas* species (Storz *et al.*, 2005). CsrA (carbon storage regulatory protein) is an RNA-binding protein in *E. coli*, which inhibits translation of mRNAs encoding proteins involved in carbon utilization (Romeo *et al.*, 1993). Two sRNAs, CsrB and CsrC, can sequester CsrA away from its binding site on the mRNAs, thereby resulting in the upregulated expression of early stationary phase processes like glycogen synthesis/catabolism and gluconeogenesis (Weilbacher *et al.*, 2003; Babitzke and Romeo, 2007).

The RsmY and RsmZ sRNAs in *Pseudomonas* species act via the same mechanism as CsrB and CsrC in *E. coli*. These sRNAs bind several copies of RsmA proteins and, in this way, block the function of the protein RsmA, which is to repress the expression of many virulence genes (Toledo-Arana *et al.*, 2007).

### 3.2.1.3 RNAs with intrinsic enzymatic activity

The third mode of action is represented by another class of sRNAs, which have intrinsic enzymatic activities, so called ribozymes. The most prominent case is RNase P, a ribonucleoprotein consisting of RNA and protein, involved in the maturation of tRNA 5' ends (Kazantsev and Pace, 2006). Please refer to chapter 3.2.3.4.3.

## 3.2.2 Some additional examples of sRNAs involved in the control of bacterial pathogenicity

### 3.2.2.1 RNAIII in *Staphylococcus aureus*

*Staphylococcus aureus* (*S. aureus*) frequently causes hospital-acquired infections. This gram-positive bacteria utilizes a global regulator of RNA nature to control gene expression, which is part of the accessory gene regulator (*agr*) system and has been studied in great detail. The *agr* regulon is a two-component signal transduction system, which consists of *agrA* (response regulator), *agrC* (histidine kinase sensor), *agrB* (to secrete *agrD*) and *agrD* (auto inducing peptide), all transcribed into one mRNA, RNAII. A second transcript of the locus, RNAIII, acts as the effector molecule (Huntzinger, E. *et al.* 2005). RNAIII (512 nt long) acts as a regulatory molecule for the expression of several virulence factors and also functions as an mRNA, encoding the small peptide  $\delta$ -hemolysin (*hld*) (Johansson and Cossart, 2003). The 5' end of RNAIII binds to the *hla* mRNA, which encodes  $\alpha$ -hemolysin, and through the binding enables translation of the *hla* mRNA. More precisely, RNAIII directly competes with an inhibitory secondary structure of the *hla* mRNA that prevents the access of ribosomes to the RBS (Romby *et al.*, 2006). The 3' end of RNAIII has also a regulatory function. It represses the expression of the *spa* gene (encoding protein A) and promotes inhibition of translation and degradation of the *spa* mRNA via an antisense mechanism that recruits RNase III. Another target of RNAIII is the *rot* (repressor of toxins) mRNA encoding a transcription factor protein (Rot), which regulates the expression of exoproteins. Translation of the *rot* mRNA is repressed by two loop-loop interactions (Geisinger *et al.*, 2006; Boisset *et al.*, 2007).

### 3.2.2.2 Qrr1-Qrr4 in *Vibrio cholerae*

*Vibrio cholerae* (*V. cholerae*) expresses the cholera toxin that causes the fatal cholera disease. The main regulators of virulence and biofilm formation of *V. cholerae* are quorum-sensing systems (QSS). One of these QQS triggers the expression of four

sRNAs (Qrr1, Qrr2, Qrr3 and Qrr4), which have very similar sequences and structures, but their sequences differ in their promoter regions, which indicates a possible differential expression regulation (Tu and Bassler, 2007). The four sRNAs target, with the help of the Hfq chaperone, the *hapR* mRNA, encoding the main regulator of many virulence genes (Toledo-Arana *et al.*, 2007). This regulation leads to a negative feedback loop in which HapR activates the expression of Qrr1-4, which in turn indirectly down-regulate HapR expression (Svenningsen *et al.*, 2008).

### **3.2.2.3 RNA $\alpha$ in *Vibrio anguillarum***

*Vibrio anguillarum* (*V. anguillarum*) is a marine fish pathogen with a tightly regulated iron-uptake system. RNA $\alpha$ , encoded within the *fatB* gene, is one system to down-regulate iron uptake. RNA $\alpha$  is a 650 nt long antisense RNA, which promotes processing of the transcript of the *fat* operon and thereby reduces the stability of the *fatB* mRNA (Chen and Crosa, 1996; Johansson and Cossart, 2003).

## **3.2.3 Regulatory RNA elements**

### **3.2.3.1 Riboswitches**

Riboswitches are *cis*-acting regulatory elements located in the 5' untranslated regions (UTRs) of mRNAs. For review please refer to Winkler and Breaker (2005). They are able to directly sense metabolite molecules and modulate transcription or translation of the downstream genes accordingly. Riboswitches consist of two domains: the aptamer domain that specifically binds a metabolite and the effector domain, the conformation of which changes thus leading to further regulation. The aptamer domains are usually highly conserved. Known ligands that bind to riboswitches are amino acids, sugars, vitamins or purines (Winkler and Breaker, 2005). No riboswitch has to date been reported in *S. pyogenes* but the glycine, TPP, purine, FMN and *yybP-ykoY* riboswitches are predicted for *S. pyogenes* in the Rfam database (Table 1).



The class of TPP riboswitches, containing a so-called *THI*-element, has been reported in many bacteria and is the only riboswitch so far that has also been identified in eukaryotes, such as plants and fungi. The effector molecule is thiamin pyrophosphate (TPP) and upon binding of TPP to the *THI*-element, the expression of genes involved in thiamine biosynthesis and transport is regulated. There are different mechanisms of regulation upon TPP binding. In *E. coli* the RBS is sequestered, while in *B. subtilis* a transcription termination mechanism is employed (Mironov *et al.*, 2002; Winkler *et al.*, 2002; Ontiveros-Palacios *et al.*, 2008).

The FMN riboswitch, containing a so-called *RFN*-element, binds flavin mononucleotide (FMN) and thereby regulates the expression of riboflavin synthesis and transport genes, like the *ribDEATH* operon in *B. subtilis*. The mechanism of regulation is most likely a premature transcription termination mechanism (Winkler *et al.*, 2002; Vogl *et al.*, 2007).

The glycine riboswitch binds specifically the amino acid glycine and regulates the glycine cleavage system, which is involved in glycine catabolism. This riboswitch was first identified in *B. subtilis*, where it controls the *gcvT* operon. In the absence of glycine the transcription of the *gcvT* operon is prematurely terminated, while the presence of glycine leads to full-length transcript and gene activation (Mandal *et al.*, 2004; Phan and Schumann, 2007).

Purine riboswitches bind guanine or adenine and were first described in *B. subtilis*, where they regulate expression of purine transport and biosynthesis genes. In this case, sensing of high intracellular purine concentrations results in the repression of the *xpt-pbuX* operon by a transcription termination mechanism (Mandal *et al.*, 2003; Mandal and Breaker, 2004).

The *yybP-ykoY* RNA element was initially identified through a bioinformatical approach. It is associated with genes encoding membrane proteins or a cation transporter and so far no ligands of this putative riboswitch are known (Barrick *et al.*, 2004).

**Table 1.** Riboswitches predicted in the Rfam database for *S. pyogenes*, which have already been described in other bacteria (adapted from (Winkler and Breaker, 2005)).

Riboswitch	Ligand	Associated gene	Organisms
RFN element (FMN)	Flavin mononucleotide	Riboflavin biosynthetic transport genes	$\alpha$ -, $\beta$ - and $\gamma$ - proteobacteria <i>Deinococcales</i> <i>Fusobacteriales</i> <i>Bacillales</i> <i>Lactobacillales</i> <i>Clostridiales/Thermoanaerobacteriales</i> <i>Actinomycetales</i> <i>Thermotogales</i>
THI element (TPP)	Thiamine pyrophosphate	Thiamine import or synthesis	$\alpha$ - $\beta$ -, $\gamma$ -, $\delta$ - and $\epsilon$ - proteobacteria <i>Deinococcales</i> <i>Bacillales</i> <i>Lactobacillales</i> <i>Clostridiales/Thermoanaerobacteriales</i> <i>Actinomycetales</i> Cyanobacteria <i>Thermotogales</i> <i>Chlorobiales</i> <i>Bacteroidales</i> <i>Fusobacteriales</i> <i>Spirochaetales</i> <i>Thermotogales</i> Archaea Eukaryotes
Glycine (gcvT)	Glycine	Glycine catabolism	$\alpha$ -, $\beta$ - and $\gamma$ - proteobacteria <i>Bacillales</i> <i>Lactobacillales</i> <i>Clostridiales/Thermoanaerobacteriales</i> <i>Actinomycetales</i>
Purine	Purine bases	Purine synthesis and transport	$\gamma$ - and $\delta$ - proteobacteria <i>Bacillales</i> <i>Lactobacillales</i> <i>Clostridiales/Thermoanaerobacteriales</i> <i>Fusobacteriales</i>
YybP-ykoY	unknown	Membrane proteins/cation transporter	$\alpha$ -, $\beta$ - and $\gamma$ - proteobacteria <i>Bacillales</i> <i>Clostridiales</i> Cyanobacteria Actinobacteria

References mentioned by Winkler and Breaker (2005): (Barrick *et al.*, 2004; Gelfand *et al.*, 1999; Griffiths-Jones *et al.*, 2005; Mandal *et al.*, 2003; Nahvi *et al.*, 2004; Rodionov *et al.*, 2002, 2003; Sudarsan *et al.*, 2003; Vitreschak *et al.*, 2002, 2003, 2004; Winkler *et al.*, 2003).

### 3.2.3.2 T boxes

T boxes are common regulatory elements with highly conserved structural elements located in the 5' UTRs of specific mRNAs in gram-positive bacteria. T box RNAs interact with charged and uncharged tRNAs to control the expression of downstream genes that are responsible for biosynthesis or transport of specific amino acids. It is likely that the ratio between the charged and uncharged tRNAs is sensed, since the two forms of tRNAs compete for access to the T box (Wakeman *et al.*, 2007). T box regulated genes are controlled by a transcription termination mechanism and transcriptional read-through is only allowed if an anti-terminator structure is formed, which is stabilized via an interaction with the uncharged tRNA. T boxes are predicted for all gram-positive organisms, except the Mycoplasma. Experimentally verified T boxes are described in *Bacillus*, *Lactococcus lactis* and *S. aureus* (Grundy and Henkin, 2003).

### 3.2.3.3 Leader RNAs

In *E. coli*, the expression of r-proteins has been described in detail. Many ribosomal proteins (r-proteins) are encoded in operons that are regulated via a feedback mechanism. In general, one protein encoded in an operon is responsible for regulation of the expression of the whole operon. These proteins bind directly to RNA, either to rRNA in the ribosome, or when it is present in excess, also to an RNA structure (operator) in the 5' UTR of the operon mRNA, thereby regulating gene expression at the translational level. In gram-positive bacteria this regulatory mechanism does not seem to be conserved, as it was shown that the *infC* operon (encoding the translation initiation factor 3 (*infC*)), r-protein L35 (*rpmI*) and r-protein L20 (*rplT*) in *B. subtilis* is regulated through premature transcription termination (Guillier *et al.*, 2002; Merianos *et al.*, 2004; Choonee *et al.*, 2007).

### 3.2.3.4 Housekeeping RNAs

#### 3.2.3.4.1 6S RNA

6S RNA has been thoroughly described in *E. coli* (Wassarman, 2007). It is a 184 nt long RNA, which accumulates during the stationary phase of growth. Although 6S RNA was one of the first described sRNAs, its function was not known for a long time. Recently it was described that 6S RNA binds to the  $\sigma^{70}$ -RNA polymerase by mimicking the structure of open promotor DNA (please refer to the section 3.2.1.2). 6S RNA seems to mediate the response to stress in stationary phase, such as for example nutrient limitation (Wassarman and Saecker, 2006).

#### 3.2.3.4.2 tmRNA

The transfer-messenger RNA (tmRNA or SsrA) is an RNA molecule that combines typical features of transfer RNAs and messenger RNAs (Withey and Friedman, 2002). tmRNA releases stalled ribosomes from mRNAs that do not have a clear stop signal, and targets degradation of the truncated proteins. In more detail, at its 3' end, tmRNA has a tRNA-like structure, charged with an alanine residue, that enters the ribosome at the A-site. Then, the ribosome switches to translation of the short SsrA-tag encoded by tmRNA, which targets the protein for clpX-mediated degradation. This process is also referred to as SsrA-tagging (Keiler, 2007).

#### 3.2.3.4.3 RNase P

RNase P has been reported in bacteria, archaea and in eukaryotes. In bacteria, the RNA component of the enzyme catalyzes the maturation of the 5' end of tRNAs, by hydrolysing a specific phosphodiester bond of the tRNA precursor. RNase P also generates the 5' end of the t-RNA like structure of tmRNA, it processes 4.5S RNA of SPR (Peck-Miller and Altman, 1991) and cleaves some structures adopted by riboswitches (Altman *et al.*, 2005). *In vitro* the protein component is not needed for RNase P function. Therefore, RNase P can be referred to as a

ribozyme or an RNA with intrinsic enzymatic activity (Kazantsev and Pace, 2006).

#### **3.2.3.4.4 SRP**

The 'signal recognition particle' (SRP) is a highly conserved ribonucleoprotein that consists in *E. coli* of the 4.5S RNA and the GTPase Ffh (48 kDa) (Bernstein *et al.*, 1989). SRP is important for the transport of proteins to or across the plasma membrane. SRP binds to a hydrophobic signal peptide of newly synthesised proteins and to FtsY, the SRP receptor bound to the plasma membrane. In this way SRP mediates a cotranslational targeting of proteins to the plasma membrane (Batey *et al.*, 2000; Herskovits *et al.*, 2000).

## 4. Materials and Methods

---

### 4.1 Computational predictions and analysis of sRNAs

#### 4.1.1 RNA predictions using computational tools

Four different programmes, sRNAPredict2 (Livny *et al.*, 2006), QRNA (Rivas and Eddy, 2001), RNAz (Washietl *et al.*, 2005) and Alifoldz (Washietl and Hofacker, 2004), were used for the prediction of sRNAs in intergenic regions. sRNAPredict2 uses sequence conservation, rho-independent terminators and putative promoters as typical features to predict sRNAs. QRNA looks for RNA structure conservation in IGRs. RNAz and Alifoldz compare genomes of different bacteria; in this case, the genome of *S. pyogenes* was compared to *S. agalactiae* 2603V/R, *S. mutans* UA159 (AEO14133) and *S. pneumoniae* TIGR4.

The candidates to validate experimentally were chosen in the following way. First, all predictions derived from sRNAPredict2 were used. Second, IGRs that contained predictions of all three programs (QRNA, RNAz and Alifoldz) were further analyzed for rho-independent terminator predictions using TransTermHP (v2.04; <http://transterm.cbcb.umd.edu/>), for putative promoters using BPROM software ([www.softberry.com](http://www.softberry.com)), for ribosome binding site (RBS) sequences (with at least 75% similarity to the consensus sequence for RBS AGAAAGGAGG) and for putative open reading frames (ORFs). Sequences of all candidates were analyzed using Vector NTI Advance™ 10 (Invitrogen). The predictions were then compared to the Rfam database (Griffiths-Jones *et al.*, 2005), which lists 18 already predicted sRNAs in *S. pyogenes* ([http://www.sanger.ac.uk/cgi-bin/Rfam/genome\\_dist.pl](http://www.sanger.ac.uk/cgi-bin/Rfam/genome_dist.pl)). The Rfam database is a collection of multiple sequence alignments and covariance models that annotates already known non-coding sRNAs in many different bacteria.

### **4.1.2 Preparation of probes for transcriptional expression analysis**

Primers annealing to internal regions of predicted sRNAs were designed. These oligonucleotides were used to create DNA or RNA probes for Northern blot analysis.

### **4.1.3 Secondary structure predictions**

Secondary structures were predicted using two computer programs, RNAfold (Hofacker *et al.*, 1994) and Mfold (Zuker, 2003).

### **4.1.4 mRNA target prediction**

mRNA targets were predicted by Tanja Gesell and Martin Grabner (CIBIV, MFPL, University of Vienna, Medical University of Vienna, Veterinary University of Vienna, Vienna (Austria)) using the method described by Mandin *et al.* (2007).

## **4.2 Bacterial strains, growth conditions and media**

### **4.2.1 Bacterial strains**

The *S. pyogenes* strains M1 SF370, M3 MGAS315, M5 Manfredo and M49 CS101 were used in this study (Table 2). The M1 strain (EC904) was isolated from an infected wound and the M3 (EC905) strain was isolated from a clinical specimen of a patient with toxic shock syndrome in Texas in the late 1980's (<http://www.lgcpromochem-atcc.com/>).

**Table 2:** *S. pyogenes* strains used in this study.

<i>S. pyogenes</i> strain	Relevant characteristics	Source
EC904, SF370	M1 serotype wild-type	ATCC 700294
EC905, MGAS315	M3 serotype wild-type	ATCC BAA 595
RDN60, M5 Manfredo	M5 serotype wild-type	Malak Kotb
RDN57, M49 CS101	M49 serotype wild-type	Bernd Kreikemeyer

The *E. coli* strains used in this study are listed in Table 3.

**Table 3.** *E. coli* strains used in this study.

<i>E. coli</i> strain	Relevant characteristics	plasmid
EC03	DH5 $\alpha$	pUC19
EC1434	Top10	pEC214
EC1254	Top10	pEC262

## 4.2.2 Growth conditions and media

### 4.2.2.1 *S. pyogenes*

*S. pyogenes* strains were grown over night (o/n) on TSA blood agar plates (12 g TSA adjusted to 400 ml with dH<sub>2</sub>O, supplemented with 3% blood (12 ml)) at 37°C in an incubator supplemented with 5% CO<sub>2</sub>. The following day an o/n equivalent suspension was made. For this, bacteria were scraped off the plate and resuspended in 1 ml THY (12 g THB and 0.8 g yeast extract adjusted to 400 ml with dH<sub>2</sub>O). The o/n equivalent suspension was diluted in THY medium to a high density (OD<sub>620 nm</sub> = 1). The cells were then grown in a 1:1000 dilution in THY at 37°C with 5% CO<sub>2</sub>. Alternatively the cells were diluted 1:1000 from an o/n culture in 25 ml THY medium.

### 4.2.2.2 *Escherichia coli*

*E. coli* strains were grown in LB medium (4 g Bacto-tryptone/peptone, 2 g Bacto-yeast extract, 2 g NaCl adjusted with dH<sub>2</sub>O to 400 ml) at 37°C, shaking the cultures at 160 rpm.



### 4.2.2.3 Antibiotics

When required, media for bacterial growth were supplemented with antibiotics at the indicated concentrations (Table 4).

**Table 4.** Concentration of used antibiotics.

Antibiotic	<i>E. coli</i>	<i>S. pyogenes</i>
Ampicillin (100 mg/ml)	100 µg/ml	-
Kanamycin (100 mg/ml)	25 µg/ml	300 µg/ml

## 4.3. DNA preparation

### 4.3.1 Plasmid preparation from *E. coli*

Plasmid preparations were performed using the E.Z.N.A.<sup>®</sup> Plasmid Miniprep Kit I (classic line) (peqLab Biotechnologie GmbH) according to the manufacturer's instructions.

#### 4.3.1.1 Quick gram-negative plasmid preparation

This protocol was used for screening clones in *E. coli*. Plasmid DNA was prepared from 2 ml of *E. coli* o/n culture. After spinning down the cells at full speed (13000 rpm) in a table-top centrifuge, the pellet was resuspended in 50 to 100 µl of culture supernatant. 300 µl of TENS buffer (TE buffer (10 mM Tris-HCl pH 8; 1 mM EDTA pH 8) supplemented with 0.1 M NaOH and 0.5% SDS) were added and the suspension was mixed gently until lysis was observed. Proteins were precipitated by addition of 150 µl of 3 M sodium acetate (pH 5.2) and gentle mixing. The cell debris was pelleted by centrifugation for 3 min at 13000 rpm. The supernatant was mixed with 900 µl of pre-cooled (at 4°C) 96% ethanol to precipitate the plasmid DNA that was then pelleted by centrifugation for another 3 min at 13000 rpm. The pellet was washed twice with 70% ethanol and dried for 5-10 min at 70°C. The DNA was resuspended in 30 µl of sterile dH<sub>2</sub>O-RNase (495 µl sterile dH<sub>2</sub>O and 5 µl RNase (10 µg/µl)), incubated for 20 min at 37°C with shaking at 300 rpm and stored at -20°C.

## 4.4 Standard DNA manipulation procedures

### 4.4.1 Polymerase chain reaction (PCR)

PCR amplifications were performed using primers listed in Table 11 (Appendix).

#### 4.4.1.1 Taq-DNA polymerase

This polymerase was used for generating PCR fragments intended for screening procedures or probes. 50  $\mu$ l reactions were prepared on ice (28.5  $\mu$ l DEPC-treated water (1 ml DEPC (Diethylpyrocarbonat) per 1 L ddH<sub>2</sub>O was stirred o/n and autoclaved the following morning.), 5  $\mu$ l 10X PCR buffer, 4  $\mu$ l 25 mM MgCl<sub>2</sub>, 5  $\mu$ l 2.5 mM dNTPs, 1  $\mu$ l 20 pmol/ $\mu$ l primer forward, 1  $\mu$ l 20 pmol/ $\mu$ l primer reverse, 50 ng genomic DNA template, 0.5  $\mu$ l *Taq*-DNA polymerase (5 U/ $\mu$ l; Fermentas)). As negative control one PCR-reaction was prepared without DNA and as positive control primers were used that did already work well in previous experiments (OLEC182+183, specific to the *fasX* RNA locus). A standard cycling program was used: 94°C for 3 min; 30 cycles: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min; 72°C for 7 min.

#### 4.4.1.2 Phusion polymerase

This polymerase was used for generating PCR fragments intended for cloning and sequencing reactions. 50  $\mu$ l reactions were prepared on ice following the manufacturer's instructions (Finnzyme). As cycling conditions, the following program was used: 98°C for 3 min; 30 cycles: 98°C for 30 sec, 58°C for 30 sec, 72°C for 30-60 sec; 72°C for 7 min.

### 4.4.2 Restriction digestions

For control digestions and cloning steps, 20  $\mu$ l and 70  $\mu$ l reactions were set up (Table 5) and incubated for at least 1 h and 3 h, respectively, at 37°C.

**Table 5.** Restriction digestions.

	20 $\mu$ l reaction	70 $\mu$ l reaction
DNA	300 ng	3000 ng
10X Enzyme buffer	2 $\mu$ l	7 $\mu$ l
Enzyme	0.5 $\mu$ l	3 $\mu$ l
dH <sub>2</sub> O	To 20 $\mu$ l	To 70 $\mu$ l

#### 4.4.3 Purification of DNA fragments

Purification of PCR products or digestion reactions was done using the QIAquick PCR purification kit according to the protocol supplied by the manufacturer (QIAGEN). The purified PCR products were eluted from the column using 30 - 50  $\mu$ l of DEPC-treated water.

#### 4.4.4 Agarose gel electrophoresis

To analyse PCR products, digestion reactions, DNA or RNA, sample aliquots were applied to 1X TAE-1% agarose gels (1 g agarose, 100 ml 1X TAE (dilution of 50X TAE buffer (242 g Tris base, 57.1 ml Glacial acetic acid, 100 ml 0.5 M EDTA pH 8.0, adjusted to 1 L ddH<sub>2</sub>O)), 4  $\mu$ l ethidium bromide [10 mg/ml]). The gels were run at 90 V for 20 min and 45 min, at which time pictures were taken.

#### 4.4.5 General cloning steps

Plasmids were prepared using the Peqlab Plasmid Miniprep Kit as described above (see 4.3.1). PCR amplifications of DNA fragments that were used as inserts were done with Phusion polymerase as described above (see 4.4.1.1). All digestion reactions of purified plasmids and DNA fragments were performed as described above (see 4.4.2). Digested vectors were treated with CIAP (calf intestinal alkaline phosphatase, Fermentas) at 37°C for 1 h to avoid self-ligation during cloning. Ligation reactions using T4 DNA ligase (Fermentas) in 1X T4 DNA ligase buffer were done in a total volume of 20  $\mu$ l and incubated o/n at 16°C or for 3 h at 22°C with shaking at 300 rpm. 10  $\mu$ l of the ligations were used for

heat-shock transformation into CaCl<sub>2</sub>-competent *E. coli* TOP10, which were used as host (see 4.4.6 and 4.4.7). Clones were selected on LB agar plates containing either ampicillin or kanamycin. Then plasmids were isolated from the selected clones and the plasmid constructs were checked with digestion reactions using the appropriate restriction enzymes (see 3.8 and 3.9).

#### **4.4.6 Preparation of competent *E. coli* cells**

An *E. coli* o/n culture was prepared in LB medium shaking at 160 rpm and 37°C. The bacteria were subcultured 1:100 in LB medium shaking at 160 rpm and 37°C until an OD<sub>620 nm</sub> of 0.5 was reached. The culture was transferred into precooled centrifuging tubes and incubated for 10 min on ice. The cells were pelleted by centrifugation for 10 min at 4500 rpm and 4°C. The supernatant was discarded. The cells were resuspended in 100 mM CaCl<sub>2</sub>, incubated for 1 h on ice and pelleted again by centrifugation for 10 min at 4500 rpm and 4°C. The pellet was resuspended in 4 ml 100 mM CaCl<sub>2</sub> buffer containing 20% glycerol and incubated for 15-60 min on ice. The cells were aliquoted into 300 µl, frozen in liquid nitrogen and stored at -80°C.

#### **4.4.7 Heat-shock transformation of competent *E. coli* cells**

100 µl competent cells (see 4.4.6) per transformation were thawed on ice, mixed gently with 10 µl of ligation product and incubated for 30 min on ice. The cells were transferred for 2 min to 42°C and then immediately chilled on ice for 5 min. 900 µl LB medium were added and incubated at 37°C shaking at 160 rpm for up to 50 min if the selection was for ampicillin resistance and for 1 h if the selection was for kanamycin resistance.

#### **4.4.8 Measurement of DNA concentration**

DNA concentrations were measured with the NanoDrop™ 1000 Spectrophotometer (Peqlab) as described in the manufacturer's instructions.

## 4.5 Primer extension

### 4.5.1 PCR product sequencing

The USB sequenase version 2.0 PCR product sequencing kit (no.: 70170) was used and sequencing was done following the manufacturer's instructions.

### 4.5.2 $\gamma$ -<sup>32</sup>P ATP Labelling of primers

To label the primer, 4  $\mu$ l of DEPC-treated H<sub>2</sub>O, 1  $\mu$ l primer [10 pmol], 1  $\mu$ l Fermentas T4 PNK 10X buffer A, 10 U T4 Polynucleotidkinase (Fermentas) and 3  $\mu$ l  $\gamma$ -<sup>32</sup>P ATP (3000 Ci/mmol) were mixed and incubated for 30 min at 37°C. The primers were denaturated for 2 min to 90°C and the samples were collected at the bottom of the tubes by centrifugation. 90  $\mu$ l DEPC-treated H<sub>2</sub>O were added to each sample.

### 4.5.3 Primer extension

RNA templates for primer extensions were obtained by total RNA extractions from *S. pyogenes* M1 strain. These samples needed to be DNA free and were therefore treated with DNase I and further checked for absence of DNA using PCR analysis.

#### 4.5.3.1 Primer extension using Promega AMV Reverse Transcriptase

For primer extension reactions, 2  $\mu$ l 5X reaction buffer, 8  $\mu$ l RNA [1  $\mu$ g/ $\mu$ l] and 1  $\mu$ l of the labelled primer were assembled and heated to 90°C for 5 min. The tubes were placed on ice for 5 min. 2.6  $\mu$ l DEPC-treated H<sub>2</sub>O, 2  $\mu$ l 5X reaction buffer, 1.4  $\mu$ l sodium pyrophosphate, 2  $\mu$ l dNTPs and 1  $\mu$ l AMV-Reverse transcriptase (10 U/ $\mu$ l) were mixed. 9  $\mu$ l of the mix were added to the tubes with the labelled extension primer and the reaction was incubated for 30 min at 42°C. After this incubation time, 20  $\mu$ l of 2X RNA loading dye were added and the samples stored at -20°C till further use.

#### 4.5.3.2 Primer extension using Invitrogen ThermoScript Reverse Transcriptase

5–10  $\mu\text{g}$  RNA, 2  $\mu\text{l}$  10 mM dNTPs and 1  $\mu\text{l}$  labelled primer were mixed and adjusted with DEPC-treated  $\text{H}_2\text{O}$  to 11  $\mu\text{l}$ . The sample was heated to  $90^\circ\text{C}$  for 5 min, placed on ice for 5 min and collected at the bottom of the tube by centrifugation. 3.3  $\mu\text{l}$  DEPC-treated water, 4  $\mu\text{l}$  5X reaction buffer and 1  $\mu\text{l}$  DTT were pre-warmed to  $37^\circ\text{C}$ , mixed in a fresh tube and 0.7  $\mu\text{l}$  ThermoScript reverse transcriptase were added. 9  $\mu\text{l}$  of the reaction mix were transferred to the tube with the labelled extension primer and mixed gently by pipetting. The sample was incubated at three different temperatures,  $55^\circ\text{C}$ ,  $60^\circ\text{C}$  and  $65^\circ\text{C}$ , for 20 min each. To inactivate the enzyme the tube was heated for 5 min to  $85^\circ\text{C}$ . 20  $\mu\text{l}$  of 2X RNA loading dye were added and the sample and stored at  $-20^\circ\text{C}$  until further use.

#### 4.5.4 Sequencing gel

Prior to pouring, the IPC (integral plate chamber) of the gel apparatus (BioRad Sequi Gen) was coated with 1 ml Dichlordimethylsilan (Merck, L813714629) and the apparatus was assembled. A 100 ml 8% PAA solution was prepared, poured with a syringe as described in the manufacturer's instructions and left to set for 1 h. The comb wells were washed with 1X TBE buffer. 3  $\mu\text{l}$  2X RNA loading dye were applied to three slots to check if the gel was running straight. The gel was pre-run for 30 min to 1 h at 90 W, 2.4 kV and 150 mA until it had reached a temperature of  $50^\circ\text{C}$ . The samples were denatured at  $95^\circ\text{C}$  for 5 min and transferred on ice. 3  $\mu\text{l}$  of each sequencing reaction were loaded in the order T, G, C and A. An aliquot of the primer extension reaction (15  $\mu\text{l}$ ) and a 1:10 dilution of the labelled primer (5  $\mu\text{l}$ ) were loaded. The gel was run for 2-4 h (90 W, 2.4 kV, 150 mA), depending on the expected size of the primer extension product. The gel was dried for 1 h at  $75^\circ\text{C}$  under vacuum and exposed in a phosphorimager cassette (Kodak) for o/n.

## 4.6 RNA techniques

### 4.6.1 Preparation of total RNA from *S. pyogenes*

Bacteria were streaked from storage at  $-80^{\circ}\text{C}$  onto TSA blood agar plates and grown o/n in an incubator at  $37^{\circ}\text{C}$ , 5%  $\text{CO}_2$ . Cells were scraped off the plate to make an o/n suspension in 1 ml THY medium with an  $\text{OD}_{620\text{ nm}}$  of  $\sim 1$ . The o/n suspension was diluted 1:1000 in 200 ml THY medium and the bacteria were grown at  $37^{\circ}\text{C}$ , 5%  $\text{CO}_2$ , not shaking. At defined time points (TP), when the  $\text{OD}_{620\text{ nm}}$  of 0.08 (TP1, early-log phase), 0.2 (TP2, early-mid-log phase), 0.32 (TP3, mid-log phase), 0.44 (TP4, late-mid-log phase) and 0.56 (TP5, late-log phase) were reached, the cells were collected for RNA extraction. TP6 and TP7 were harvested 1 h after TP5 and after o/n growth respectively representing early and late stationary phase. The following amounts of culture were harvested: 40 ml at TP1, 20 ml at TP2, 10 ml at TP3, TP4, TP5, TP6 and TP7. The bacterial culture was mixed 1:1 with ice cold 1:1 acetone-ethanol and immediately frozen at  $-80^{\circ}\text{C}$ . The harvested cells were centrifuged for 1 min at 13000 rpm and  $4^{\circ}\text{C}$ , the pellet was washed once with 300  $\mu\text{l}$  TE/25% sucrose (2.5 ml 1 M Tris-HCl pH 8, 1 ml 0.5 M EDTA pH 8, 12.5 g sucrose, adjusted to 50 ml with DEPC-treated water) and resuspended in 300  $\mu\text{l}$  TE/25% sucrose. 150  $\mu\text{l}$  cells were used for one RNA preparation. For this, the cells were pelleted by centrifugation and resuspended in 100  $\mu\text{l}$  lysis buffer (1 ml 1 M Tris-HCl pH 8, 5 ml 0.5 M EDTA pH 8, 10 g sucrose, adjusted to 50 ml with DEPC-treated water. Just before use, 5  $\mu\text{l}$  of lysozyme (50 mg/ml) and 5  $\mu\text{l}$  of mutanolysin (10 U/ $\mu\text{l}$ ) were added per 100  $\mu\text{l}$  lysis buffer). The samples were incubated on ice for 5 min and then 150  $\mu\text{l}$  of lysis executioner (850  $\mu\text{l}$  DEPC-treated water, 50  $\mu\text{l}$  Proteinase K (20 mg/ml), 100  $\mu\text{l}$  20% SDS) were added. The samples were heated for 1 min to  $95^{\circ}\text{C}$ , cooled down at RT. When lysis was completed, 750  $\mu\text{l}$  of TRI Reagent (Sigma) were added. The samples were incubated for 5 min at RT, then 200  $\mu\text{l}$  of chloroform were added. The samples were shaken vigorously for 15 sec and incubated for 10 min at RT. The tubes were centrifuged for 10 min at  $4^{\circ}\text{C}$  (10000 rpm), the aqueous phase was

transferred to a clean tube and 500  $\mu$ l of isopropyl alcohol were added. The samples were inverted, incubated for 10 min at RT and then the RNA was pelleted by centrifugation for 10 min at 4°C (10000 rpm). The RNA pellet was washed once with 1 ml of 70% ethanol, air-dried and resuspended in 30  $\mu$ l of DEPC-treated water.

The integrity of RNA was assessed by loading 2  $\mu$ l of the samples with 5  $\mu$ l 2X RNA loading buffer and 3  $\mu$ l DEPC-treated water on a 1X TAE-1% agarose gel. The RNA concentration was measured using the NanoDrop™ 1000 Spectrophotometer (PqLab). Before the RNA was used in experiments, the DNA in the sample had to be eliminated. Therefore, the sample was brought to a concentration of 1  $\mu$ g/ $\mu$ l with DEPC-treated water, denatured for 10 min at 65°C, kept for 5 min on ice and treated with 3  $\mu$ l DNase I (1 U/ $\mu$ l Fermentas) for 30 min at 37°C. The reaction was stopped at -20°C. If the RNA was used for primer extension experiments, the samples were checked for residual DNA content by PCR using *fasX* primers OLEC182+183 (Table 11).

#### **4.6.2 PAA gel electrophoresis**

8% PAA gels (Table 6) were used for electrophoresis of total RNA extracts. After gel preparation, the gel wells were rinsed with 1X TBE. 10  $\mu$ l of the RNA samples were mixed with 10  $\mu$ l 2X RNA loading buffer (8.4 ml formamide, 25 mg xylene cyanol FF, 25 mg bromphenol blue, 10  $\mu$ l 0.5 M EDTA, 12.5  $\mu$ l 20% SDS, 1.6 ml DEPC-treated water), denatured for 10 min at 65°C and kept on ice until loading. The whole sample (20  $\mu$ l volume) was loaded into one slot, 5  $\mu$ l of RNA low range ladder (Fermentas) were used as marker. The gels were run at 15 mA per gel for 1 h 10 min. Then the gels were stained with ethidium bromide in 1X TBE for 10 min and a picture was taken. Before the gels were used for RNA transfer to membranes they were washed 10 min with 1X TBE buffer (10 min).



**Table 6.** PAA gel.

	<b>8% PAA-Gel (15 ml)</b>
<b>25% Acryl-Bisacrylamide (19:1) – 8 M Urea</b> (625 ml Acrylamide/Bisacrylamide solution 40% (1:19), 480.48 g Urea adjusted to 1 l)	4.8 $\mu$ l
<b>10X TBE</b> (54 g Tris-base, 27.5 g Boric acid, 20 ml 0.5 M EDTA adjusted to 1 l)	1.5 $\mu$ l
<b>8 M Urea</b> (480.48 g Urea per 1 l DEPC-treated water)	8.7 $\mu$ l
<b>10% APS</b> (1 g Ammonium Persulfate in 10 ml ddH <sub>2</sub> O)	110 $\mu$ l
<b>Temed</b> (Tetramethylethylenediamine)	15 $\mu$ l

### 4.6.3 Northern blot analysis using PAA gels

The blots were set up in a semi-dry blotting apparatus (BioRad Transblot, 3X Whatman paper, Hybond-N<sup>+</sup> membrane (GE Healthcare), gel, 3X Whatman paper). As transfer buffer 1X TBE was used and the transfer was done for 1 h at 12 V. After blotting, the slots and the ladder bands were marked on the membrane using a membrane-marker pen. The RNA was cross-linked to the membrane twice using the 'autocrosslink' function of a strata-linker, with a rinse in 6X SSC in between the crosslinking steps. The membranes were wrapped in cling film and stored at 4°C.

For Northern blot hybridization the membrane was prehybridized with prehybridization solution (30 ml Hybridization solution (36 ml Denhardt 50X (1 g Ficoll 400, 1 g Polyvinylpyrrolidone, 1 g Bovine Serum Albumin adjusted to 100 ml with DEPC-treated water), 225 ml 20X SSC (350.6 g NaCl, 177.4 g Sodium citrate, adjusted to 2 L with DEPC-treated water), 18 ml 0.5 M EDTA pH 8 (74.45 g Ethylenediaminetetraacetic acid adjusted to 400 ml with DEPC-treated water), 9 ml 20% SDS (80 g Sodium dodecyl sulfate adjusted to 400 ml DEPC-treated water), adjusted to 900 ml with DEPC-treated water), 300  $\mu$ l denatured Salmon sperm DNA) for 2 h at 55°C and the probe was hybridized to the membrane o/n in hybridization solution.

Then, the membrane was first shortly rinsed and then washed for 30 min with washing buffer I (25 ml 20X SSC, adjusted to 1 L with DEPC-treated water). This was followed with a 30 min wash with washing buffer II (50 ml 20X SSC, 6.25 ml 20% SDS, adjusted to 1 L with DEPC-treated water) and a subsequent 30 min washing step in buffer III (25 ml 20X SSC, 6.25 ml 20% SDS, adjusted to 1 L with DEPC-treated water). The membranes were exposed to a phosphorimager cassette and subsequently to photo films (Agfa). As a control, the membranes were hybridized with a probe specific to either the 5S or 16S rRNA, to verify that the individual RNA samples were equally loaded.

#### **4.6.4 Northern Blot using agarose gels**

1.2 g agarose were melted in 100  $\mu$ l DEPC-treated water. When the solution was cooled to about 60°C, 12 ml 10X MOPS buffer (41.86 g MOPS, 6.8 g Sodium acetate-trihydrate, 20 ml 0.5 M EDTA pH 8 adjusted to 1 L with DEPC-treated water; filter-sterilized and stored in the dark.), 19.4 ml formaldehyde and 1  $\mu$ l ethidium bromide were added and then the gel was poured. 25-30  $\mu$ g of RNA were mixed with an equal amount of 2X RNA loading dye and denatured for 15 min to 65 °C, then chilled on ice. The gel was pre-run for 10 min at 90 V (Running buffer (100 ml MOPS, 20 ml formaldehyde adjusted to 1 L with DEPC-treated water); stored in the dark). The samples were loaded and the gel was run for 10 min at 90 V, then for 80 min at 80 V. After 40 min at 80 V, the running buffer was mixed to compensate for the pH gradient that was built up during the gel run. The gel was rinsed 3 times for 10 min with DEPC-treated water to remove the formaldehyde. For RNA transfer 3 pieces of Whatman paper were pre-wet with 20X SSC. The paper was cut to a size that allowed continuous contact with the transfer buffer. The gel was placed upside down on the Whatman paper and Parafilm was used to cover the edges of the gel, in order to prevent evaporation of the buffer. The membrane was placed on top of the gel and 3 pieces of Whatman paper, pre-wet with 20X SSC, followed by paper towels, a plate and weights were placed on top of the gel. Transfer was allowed to take place o/n.

The following day, the blotting stack was disassembled, the membranes were checked under UV-light and the slots and the ladder bands were marked using a membrane-marker pen. The RNA was cross-linked to the membrane and stored as described above (see 4.6.3).

#### 4.6.5 Labelling (Random Priming) of probes

The DNA probes were labelled with  $\alpha$ -<sup>32</sup>P-dATP. For this, 100 ng DNA were used in a volume of 60  $\mu$ l DEPC-treated water. The DNA was denatured by boiling for 10 min at 95°C and then chilled on ice. A master mix (per sample 10  $\mu$ l Oligo labelling buffer (OLB: 373  $\mu$ l solution A (1 ml Solution 0 (7.57 g Tris base, 1.27 g MgCl<sub>2</sub>, adjusted to 50 ml), 18  $\mu$ l  $\beta$ -mercaptoethanol, 5  $\mu$ l 100 mM dCTP, 5  $\mu$ l 100 mM dGTP, 5  $\mu$ l 100 mM dTTP), 933  $\mu$ l solution B (28.83 g HEPES adjusted to 50 ml) and 560  $\mu$ l Solution C (Hexadeoxyribonucleotides pd(N)<sub>6</sub> sodium salt from Amersham in TE (90 OD U/ml)) were mixed), 2  $\mu$ l BSA (10 mg/ml) and 1  $\mu$ l Klenow Fragment (10 U/ $\mu$ l)) was prepared. The master mix was added to the tubes containing the DNA and 1-3  $\mu$ l of  $\alpha$ -<sup>32</sup>P dATP (3000 Ci/mmol) were added. The tubes were incubated for 1 h at 37°C, purified with MicroSpin G-50 Columns (Amersham Pharmacia Biotech) according to the manufacturer's instructions and the labelled probe was mixed with 30 ml hybridization buffer. After hybridization, the probes were stored at -20°C for further use.

#### 4.6.6 Determination of RNA stability

For rifampicin treatment, *S. pyogenes* M1 strain was grown until the desired density in THY medium and then rifampicin (stock solution: 25 mg/ml rifampicin (Sigma) in methanol) was added to the bacterial culture at a final concentration of 250  $\mu$ g/ml. Cells were harvested at the defined timepoints, before (-1), immediately after (0 min) and 2.5 min, 5 min, 7.5 min, 10 min, 15 min and 30 min after addition of the antibiotic, for RNA extraction. To check the sensitivity of the cells to the antibiotic, cells were plated on TSA/rifampicin plates (250  $\mu$ g/ml

rifampicin) and colonies were counted after 24 h incubation at 37°C, 5%CO<sub>2</sub>. Total RNA was prepared from the collected cells and used for Northern blot analysis. The intensity of the bands was calculated relative to the signal of the 5S rRNA that was used as loading control.

## 4.7 StreptoTag

### 4.7.1 *In vitro* transcription of tagged RNA

The DNA template for *in vitro* transcription of the hybrid RNA was obtained by a two-step PCR with Phusion™ High-Fidelity DNA polymerase (Finnzymes). The *in vitro* transcription was done with a Epicentre Biotechnologies AmpliScribe™ T7-Flash™ Transcription kit following the manufacturer's instructions. The *in vitro* transcribed RNA was loaded on a 5% PAA gel, the band of the expected size was excised from the gel and then eluted o/n at 4°C, shaking at 1000 rpm in 0.3 M NaAc (pH 5.5). The following day the RNA was precipitated in 2.5 volumes 96% EtOH at -20°C for 5-6 h. The RNA was pelleted by centrifugation for 15 min, 13200 rpm at 4°C, the pellet was dried and resuspended in 50 µl DEPC-treated water.

### 4.7.2 Total cell lysates

In principle, total cell lysates of *S. pyogenes* should be prepared from cultures grown to timepoints, at which a phenotype of the sRNA of interest is expected. 50 ml of harvested cells were harvested by brief centrifugation for less than 1 min at ~6000 rpm, washed once with 1X PBS and then resuspended in 1 ml 1X PBS. The cells were sonicated for 4X 2 min cycles on ice. The cell lysate was centrifuged for 30 min at 13200 rpm at 4°C and the clear supernatant was transferred into a fresh tube. The protein concentration was determined by a Bradford protein assay. For this, 5 µl of total cell lysate were well mixed with 495 µl of Bradford reagent (BioRad), incubated for 5-30 min and the protein concentration was measured at 595 nm.

The columns were prepared and used as described (Windbichler and Schroeder, 2006). Samples of the flow through, the last washing steps and the eluate were collected and used for SDS-PAGE analysis.

### 4.7.3 SDS-PAGE analysis

Proteins of the different collected samples were separated on SDS-PAGE (Table 7) at 25 mA using 1X SDS running buffer (30.2 g Tris-base, 187.6 g Glycine, 50 ml 20% SDS, adjusted with dH<sub>2</sub>O to 1 L). The samples of the washing steps and the flow through were precipitated with 0.1 volume of 100% TCA and directly resuspended in 6X SDS loading dye. 5 µl of the flow-through were directly mixed with 6X SDS loading dye. The samples were heated to 95°C for 5 min and then put on ice before loading.

**Table 7.** SDS-PAGE

	Resolving gel (15%)	Stacking gel (3.6%)
	10 ml	3 ml
Solution A (30% Acrylamide)	4 ml	0.36 ml
Solution B (1.5 M Tris-HCl pH 8.8, 0.4% SDS)	2.5 ml	-
Solution C (0.5 M Tris-HCl pH 6.8, 0.4% SDS)	-	0.75 ml
dH <sub>2</sub> O	2.43 ml	1.89 ml
10% APS	66 µl	30 µl
TEMED	3.3 µl	3.0 µl

The gels were stained following the colloidal Coomassie staining procedure. After fixation for 30 min in a 40% ethanol, 10% acetic acid solution, gels were washed with dH<sub>2</sub>O 3X for 10 min. Staining was done o/n in the staining solution (50 g ammonium sulfate, 6 ml 85% phosphoric acid, 490 ml dH<sub>2</sub>O and 10 ml colloidal brilliant blue stock (5 g Coomassie brilliant blue G-250 in 100 ml H<sub>2</sub>O)) freshly diluted 1:5 with methanol. After staining, gels were washed with dH<sub>2</sub>O,

wrapped in plastic and the result was recorded by scanning (Scanner: Epson perfection 2400 Photo).

## 5. Aims of the study

---

### **Project 1: Analysis of novel small non-coding RNAs in *S. pyogenes***

The aim of this study was to identify and characterize novel small non-coding RNAs in *S. pyogenes* likely to play important roles in metabolic and pathogenicity-linked processes. For this purpose, a large-scale computational approach was used, which allowed the identification of intergenic regions (IGRs) likely to encode sRNAs in the *S. pyogenes* genome. Selected predictions were experimentally validated using Northern blot analysis. The 5' and 3' extremities of the sRNAs were mapped and their stability was tested in rifampicin experiments. *In silico* predictions of the secondary structure of some of the sRNA candidates were done. A computational program was also used to identify putative mRNA targets of most promising sRNA candidates. Two sRNA candidates, SpyRNA014 and SpyRNA049, were studied further in this work.

### **Project 2: Novel pleuromutilin derivatives and their interactions with *Escherichia coli* and *Staphylococcus aureus* ribosomes**

Pleuromutilin derivatives are protein synthesis inhibitors, which have been successfully used in veterinary medicine. Since an increasing number of bacterial pathogens acquired antimicrobial resistance, these compounds became recently of interest for human treatment. In this study, novel chemically designed pleuromutilin derivatives for human treatment were compared to conventional pleuromutilins in chemical footprinting experiments to determine the binding sites of the compounds to the 23S rRNA target of *E. coli* and *S. aureus*. Additionally, *in vitro* antibacterial activities against several clinical bacterial isolates were analyzed. My contribution to this work was to perform several

footprinting experiments of pleuromutilins with the ribosomes of *E. coli* and *S. aureus* (Appendix B).



## 6. Result

---

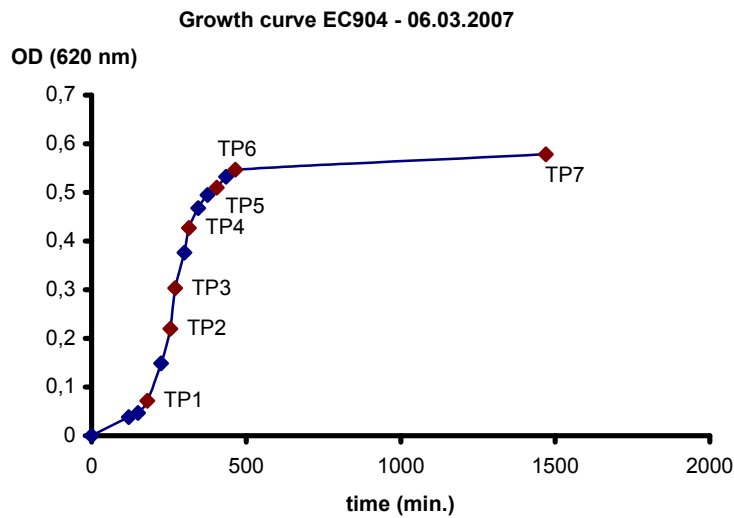
### 6.1 Selection of sRNA candidates for experimental validation and analysis (preliminary data)

Four different bioinformatic programmes, sRNAPredict2, QRNA, RNAz and Alifoldz, were used to predict putative sRNAs the IGRs of the *S. pyogenes* M1 strain SF370 genome. The predicted candidates were also compared to the Rfam database, a database listing predictions for conserved, already known RNAs previously described in other bacteria (Griffiths-Jones *et al.*, 2005). The predictions were analyzed by Maria Eckert as described in Materials and Methods, resulting in a total of 253 predictions derived from the 4 different programs. Some predictions were located in the same IGRs and were therefore merged into one putative sRNA locus. Overall 180 individual loci, predicted to encode a sRNA were identified in IGRs. 90 (Table 12) of them were chosen for further experimental analysis.

### 6.2 Experimental validation of expression of sRNA candidates using a Northern blot screen

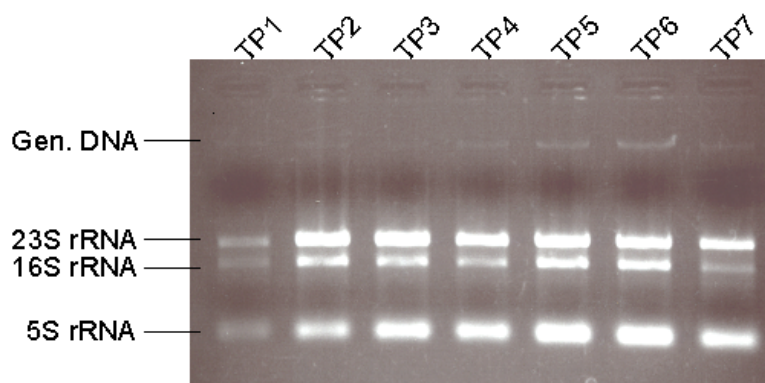
First, predicted sRNAs in IGRs were analyzed for their expression in *S. pyogenes* cultured in rich THY medium throughout the entire growth curve by Northern blot analysis.

The normal growth curve of *S. pyogenes* can be divided into three phases: lag phase, log or exponential phase and stationary phase. The cells of all four, M1, M3, M5 and M49 *S. pyogenes* clinical isolates, used in the study, were grown in normal conditions and harvested at seven different time points throughout growth (Fig. 5).



**Fig. 5:** A representative growth curve of the *S. pyogenes* strain M1 SF370 (EC904). Bacteria were grown at 37°C with 5% CO<sub>2</sub> in THY medium.

From the bacterial cells harvested at the different time points (Fig. 5), total RNA was extracted. To ensure integrity of the RNA, the preparations were analyzed by agarose gel electrophoresis (Fig. 6). Bands of 23S rRNA, 16S rRNA and 5S rRNA are clearly distinguishable since the rRNAs represent an important portion of the RNA pool in a cell. The RNA concentration was verified using a NanoDrop™ 1000 Spectrophotometer (Peachlab).



**Fig. 6.** Agarose gel electrophoresis of total RNA preparation of a *S. pyogenes* culture grown to time points 1-7 (TP1-7) in THY medium at 37°C with 5% CO<sub>2</sub>.

For Northern blot analysis, PAA gels were used as they have good resolution to separate sRNA molecules of sizes between 100 and 500 nucleotides (nt). Northern blots using agarose gels are more suitable for analysis of larger RNA molecules. Radiolabelled PCR-generated DNA fragments specific to the sRNA candidates were used as probes to highlight the sRNA expression profiles. For each candidate, the expression profile was analyzed in 4 different M types (M1, M3, M5 and M49).

In this Northern blot screen, 90 sRNA candidates from the computational predictions were chosen for expression analysis. Positive hybridization was obtained for 28 candidates (32%) (Table 8). These 28 positive candidates included 5 novel sRNAs (SpyRNA027, SpyRNA034, SpyRNA042, SpyRNA049 and SpyRNA069), SpyRNA014 (a sRNA candidate that was already once mentioned in the literature), 6 leader RNAs (L10, L13, L19, L20, L21, PyrR), 5 riboswitches (FMN, TPP, purine, glycine and yybP-ykoY), 4 functional RNAs (housekeeping RNAs; RNaseP, 6S RNA, SRP and tmRNA) and 8 T boxes (*pheST*, *alaS*, *ileS*, *valS*, *glyQS*, *serS*, *thrS* and *trsA*).

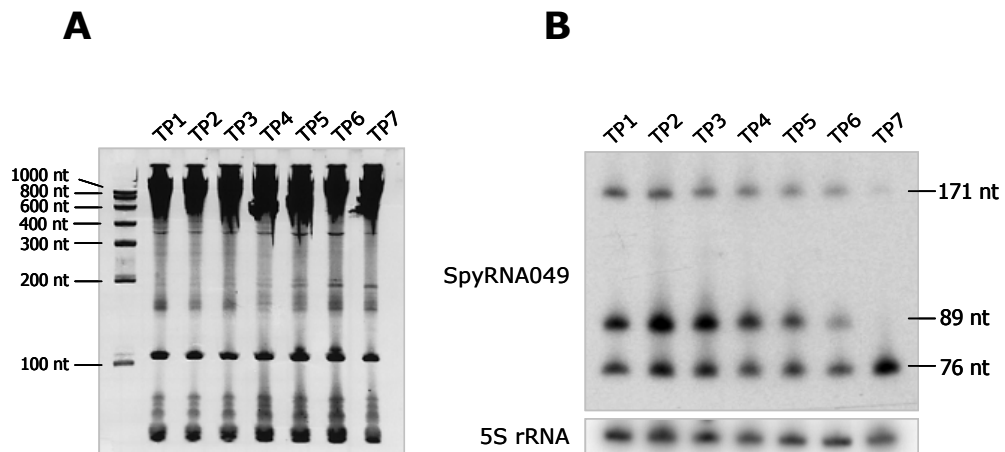
**Table 8.** List of sRNA candidates detected by Northern blot analysis and computer programs by which they were identified.

Predictions	sRNA Predict2	QRNA	RNAz	Alifoldz	Rfam	Size <sup>1</sup> (nt)	TP <sup>2</sup>
SpyRNA014	✓	✓	-	-	-	~180	1-4
SpyRNA027	✓	-	✓	✓	-	~130, 400	1-4
SpyRNA034	✓	-	-	-	-	~110	1-2
SpyRNA042	✓	✓	✓	✓	-	~195	1-5
SpyRNA049	✓	✓	-	-	-	~80, 90,170	1-6
SpyRNA069	-	✓	-	-	-	~190	1-4
L10 leader	✓	✓	✓	-	✓	~130	1-3
L13 leader	-	✓	✓	✓	✓	~900	1-5
L19 leader	✓	✓	✓	✓	✓	~500	1-4
L20 leader	✓	✓	✓	✓	✓	~100	1-4
L21 leader	-	✓	✓	-	✓	~800	1-6
yybP-ykoY	-	-	✓	✓	✓	~110	1-4
PyrR	✓	-	✓	✓	✓	~110	1-5
Glycine	✓	✓	-	-	✓	~190, 300	1-4
Purine riboswitch	✓	✓	✓	-	✓	~180	1-3
TPP riboswitch	✓	-	-	-	✓	~150	1-4
FMN riboswitch	✓	✓	✓	-	✓	~110, 200	1-4
SRP	✓	✓	✓	✓	✓	~70	1-7
tmRNA	-	✓	-	✓	✓	~400	1-6
6S RNA	-	✓	-	✓	✓	~190, 200	4-7
RNase P	✓	✓	✓	✓	✓	~400	1-5
<i>pheS</i> T box	-	-	-	✓	-	~200	1-4
<i>alaS</i> T box	-	✓	✓	-	-	~180	1-4
<i>ileS</i> T box	✓	-	✓	✓	-	~200, 220	1-4
<i>valS</i> T box	✓	-	-	✓	-	~200, 210	1-4
<i>glyQS</i> T box	-	-	✓	-	-	~220	1-5
<i>serS</i> T box	-	✓	-	-	-	~180	1-4
<i>thrS</i> T box	✓	✓	-	-	✓	~210	1-4
<i>trsA</i> T box	✓	-	-	-	✓	~250	1-4

<sup>1</sup> Size estimated from Northern blot hybridization.

<sup>2</sup> Timepoints at which the sRNA is strongly expressed (expression profile).

As an example of positive hybridization in Northern blot analysis, the expression profile of candidate SpyRNA049 is shown in Fig. 7. SpyRNA049 is a very promising sRNA candidate and therefore it was chosen as an example for all further analysis that was done in this study.



**Fig. 7.** (A) Total RNA samples loaded on a 8% PAA gel stained with ethidium bromide. This gel was used for Northern blotting. Total RNA was extracted from EC904 (M1) cells at different TPs throughout growth (refer to Fig. 5 – growth curve). (B) Northern blot hybridized with a radiolabelled SpyRNA049-specific DNA probe. A DNA probe specific for 5S rRNA was used as loading control.

SpyRNA049 is expressed in a growth-phase dependent manner. Three RNA fragments of different sizes, 170 nt, 90 nt and 80 nt (Table 8), were detected. SpyRNA049 was only observed in the M1 strain (Fig. 7), and was not detected in strains of M3, M5 and M49 types (data not shown). The sequence of the SpyRNA049 IGR is not conserved in the strains included in this study, which seems to be the reason for the lack of SpyRNA049 expression in the M types M3, M5 and M49. Alignment of the SpyRNA049 IGRs among these clinical isolates is represented in Fig 8. The IGR is only partly conserved, of which the M3 strain has the most conserved sequence compared to M5 and M49 strains.

CLUSTAL W(1.60) multiple sequence alignment

```

SF370          TTTTGCCTCCTAAAAATAAAAAGTTTAAATTAATCCATAATGAGTTTGATGATTTCAAT
MGAS2096      --TTTGCCTCCTAAAAATAAAAAGTTTAAATTAATCCATAATGAGTTTGATGATTTCAAT
MGAS5005      TTTTGCCTCCTAAAAATAAAAAGTTTAAATTAATCCATAATGAGTTTGATGATTTCAAT
MGAS6180      TTTTGCCTCCTAAAAATAAAAAGTTTAAATTAATCCATAATGAGTTTGATGATTTCAAT
MGAS9429      TTTTGCCTCCTAAAAATAAAAAGTTTAAATTAATCCATAATGAGTTTGATGATTTCAAT
MGAS10270     TTTTGCCTCCTAAAAATAAAAAGTTTAAATTAATCCATAATGAGTTTGATGATTTCAAT
MGAS10750     TTTTGCCTCCTAAAAATAAAAAGTTTAAATTAATCCATAATGAGTTTGATGATTTCAAT
SSI-1         TTTTGTCTCCTAAAAATAAAAAGTTTAAATTAATCCATAATGAGTTTGATGATTTCAAT
MGAS315       TTTTGTCTCCTAAAAATAAAAAGTTTAAATTAATCCATAATGAGTTTGATGATTTCAAT
Manfredo     -----
MGAS8232     -----
MGAS10394     -----

SF370          AATAGTTTTAATGACCTCCGAAATAGTTTAAATATGCTTTAATTTTCTTTTCAAATA
MGAS2096      AATAGTTTTAATGACCTCCGAAATAGTTTAAATATGCTTTAATTTTCTTTTCAAATA
MGAS5005      AATAGTTTTAATGACCTCCGAAATAGTTTAAATATGCTTTAATTTTCTTTTCAAATA
MGAS6180      AATAGTTTTAATGACCTCCGAAATAGTTTAAATATGCTTTAATTTTCTTTTCAAATA
MGAS9429      AATAGTTTTAATGACCTCCGAAATAGTTTAAATATGCTTTAATTTTCTTTTCAAATA
MGAS10270     AATAGTTTTAATGACCTCCGAAATAGTTTAAATATGCTTTAATTTTCTTTTCAAATA
MGAS10750     AATAGTTTTAATGACCTCCGAAATAGTTTAAATATGCTTTAATTTTCTTTTCAAATA
SSI-1         AATAGTTTTAATGACCTCCGAAATAGTTTAAATATGCTTTAATTTTCTTTTCAAATA
MGAS315       AATAGTTTTAATGACCTCCGAAATAGTTTAAATATGCTTTAATTTTCTTTTCAAATA
Manfredo     -----
MGAS8232     -----
MGAS10394     -----

SF370          TCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATCTTT
MGAS2096      TCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATCTTT
MGAS5005      TCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATCTTT
MGAS6180      TCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATCTTT
MGAS9429      TCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATCTTT
MGAS10270     TCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATCTTT
MGAS10750     TCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATCTTT
SSI-1         TCTCTTCAAAAAATATTACCTAATACTTAATAATGAATAAATATAACACAAAATCTTT
MGAS315       -----
Manfredo     -----
MGAS8232     -----
MGAS10394     -----

SF370          TAAAAAGTAGTTTATTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAA
MGAS2096      TAAAAAGTAGTTTATTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAA
MGAS5005      TAAAAAGTAGTTTATTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAA
MGAS6180      TAAAAAGTAGTTTATTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAA
MGAS9429      TAAAAAGTAGTTTATTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAA
MGAS10270     TAAAAAGTAGTTTATTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAA
MGAS10750     TAAAAAGTAGTTTATTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAA
SSI-1         TAAAAAGTAGTTTATTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAA
MGAS315       -----
Manfredo     -----
MGAS8232     -----
MGAS10394     -----

SF370          ATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAAAAGTTATAAAAATATCTTTGCGAA
MGAS2096      ATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAAAAGTTATAAAAATATCTTTGCGAA
MGAS5005      ATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAAAAGTTATAAAAATATCTTTGCGAA
MGAS6180      ATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAAAAGTTATAAAAATATCTTTGCGAA
MGAS9429      ATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAAAAGTTATAAAAATATCTTTGCGAA
MGAS10270     ATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAAAAGTTATAAAAATATCTTTGCGAA
MGAS10750     ATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAAAAGTTATAAAAATATCTTTGCGAA
SSI-1         ATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAAAAGTTATAAAAATATCTTTGCGAA
MGAS315       -----
Manfredo     -----
MGAS8232     -----
MGAS10394     -----

SF370          CCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
MGAS2096      CCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
MGAS5005      CCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
MGAS6180      CCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
MGAS9429      CCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
MGAS10270     CCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
MGAS10750     CCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
SSI-1         TTTTTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
MGAS315       TTTTTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
Manfredo     CCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
MGAS8232     CCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
MGAS10394     CCATTCAAACAGCATAGCCAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG

SF370          GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
MGAS2096      GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
MGAS5005      GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
MGAS6180      GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
MGAS9429      GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
MGAS10270     GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
MGAS10750     GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
SSI-1         GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
MGAS315       GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
Manfredo     GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
MGAS8232     GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA
MGAS10394     GCACCGAGTCGGTGCTTTTTTTT-GATACTTCTATTCTACTCTGACTGCAAAACCAAAAAA

```

```

SF370          CAAGCGCTTTCAAAACGCTTGTTTTATCATTTCAGGAAATTAATCTCTTAATCCTTTT
MGAS2096      CAAGCGCTTTCAAAACGCTTGTTTTATCATT-----
MGAS5005      CAAGCGCTTTCAAAACGCTTGTTTTATCATTTCAGGAAATTAATCTCTTAATCCTTTT
MGAS6180      CAAGCGCTTTCAAAACGCTTGTTTTATCATT-----
MGAS9429      CAAGCGCTTTCAAAACGCTTGTTTTATCATT-----
MGAS10270     CAAGCGCTTTCAAAACGCTTGTTCTATCATTTCAGGAAATTAATCTCTTAATCCTTTT
MGAS10750     CAAGCGCTTTCAAAACGCTTGTTTTATCATTTCAGGAAATTAATCTCTTAATCCTTTT
SSI-1         CAAGCGCTTTCAAAACGCTTGTTTTATCATTTCAGGAAATTAATCTCTTAATCCTTTT
MGAS315      CAAGCGCTTTCAAAACGCTTGTTTTATCATTTCAGGAAATTAATCTCTTAATCCTTTT
Manfredo     CAAGCGCTTTCAAAACGCTTGTTTTATCATTTCAGGAAATTAATCTCTTAATCCTTTT
MGAS8232     CAAGCGCTTTCAAAACGCTTGTTTTATCATTTCAGGAAATTAATCTCTTAATCCTTTT
MGAS10394    CAAGCGCTTTCAAAACGCTTGTTTTATCATTTCAGGAAATTAATCTCTTAATCCTTTT

```

**Fig. 8.** Alignment of SpyRNA049 locus (coordinates M1 SF370 NC\_002737: 854279 – 854756, complementary strand). Red: alternative 5' ends; grey: putative rho-independent terminator. *S. pyogenes* M1 is strain SF370 and MGAS5005, *S. pyogenes* M3 is strain MGAS315 and SSI-1, *S. pyogenes* M5 is strain Manfredo. *S. pyogenes* M12 is strain MGAS2096 and MGAS9429, *S. pyogenes* M28 is strain MGAS6180, *S. pyogenes* M2 is strain MGAS10270, *S. pyogenes* M4 is strain MGAS10750, *S. pyogenes* M18 is strain MGAS8232, *S. pyogenes* M6 is strain MGAS10394.

### 6.3 Mapping of 5' and 3' extremities of sRNA

To determine the exact transcription starts and stops of the sRNAs, the 5' and 3' ends were mapped. The 5' ends of candidates were determined using primer extension (Fig. 9, Table 9). The 5' and 3' ends were also mapped simultaneously using self-ligation and by 3' RACE (Maria Eckert).

**Table 9.** Coordinates of 5' and 3' ends of the sRNA candidates mapped and their exact length.

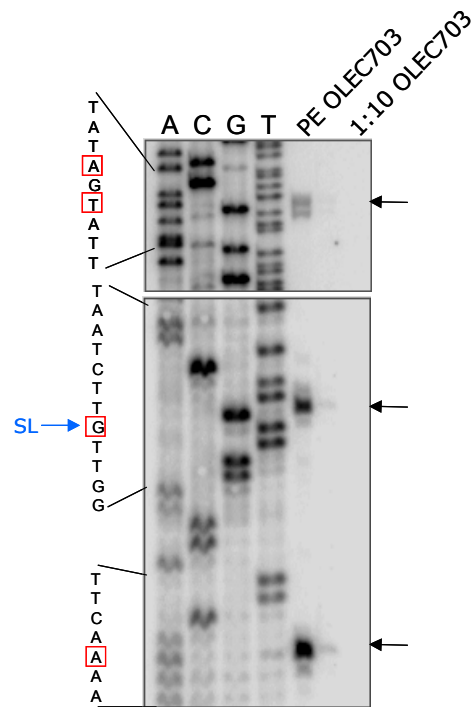
Candidate	PE	SL		length (nt)
	5' end	5' end	3' end	
SpyRNA014	1680671	1680671	1680500	172
SpyRNA027	996147	996147	995997	151
SpyRNA034	nd	1822343	1822473	131
SpyRNA042	1106555	1106554	1106379	180
SpyRNA049	854546			171
	854464	854464	854376	89
	854451			76
SpyRNA069	825970	825970	826157	188
yybP-ykoY	500055	500049	500252	104
PyrR	680207	680207	680338	132
Glycine riboswitch	1046345	1046345	1046621	277
Purine riboswitch	nd	930739	930908	180
TPP riboswitch	482963	482963	482816	148
FMN riboswitch	319190	319190	319385	196
SRP	190336	190336	190413	78
tmRNA	1065030	1065030	1065372	343
6S RNA	1663522	1663522	1663724	202
	1663529	1663530	1663724	195
Rnase P	1365555	1365555	1365139	416
SpyRNA050 (thrS)	416968	416968	417180	213
T box (tsrA)	1842269	1842269	1842016	254
		1842269	1842017	253

PE stands for primer extension. SL stands for self ligation

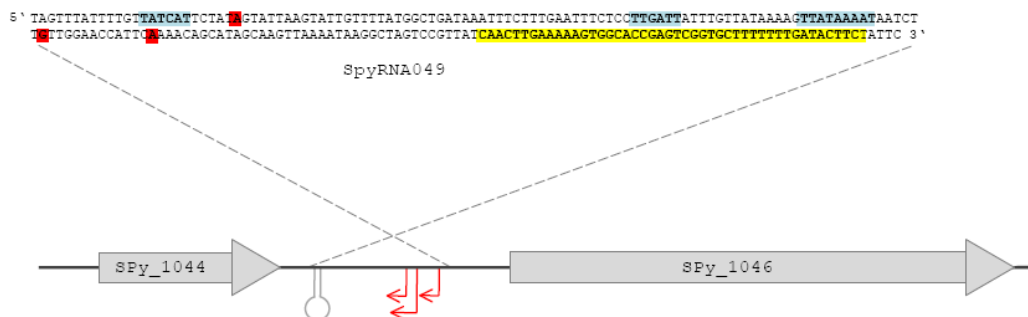
Of the 5 novel sRNAs, 3 were encoded in 5' UTRs of the downstream mRNAs (SpyRNA034, SpyRNA042 and SpyRNA069), one was encoded in 3' UTR (SpyRNA027) and one candidate was encoded independently in an IGR (SpyRNA049). As an example of mapping, primer extension and self-ligation analysis of SpyRNA049 are shown in Fig. 9 and 10, respectively. SpyRNA049 is encoded on the negative strand in the intergenic region flanked by SPy1044 and Spy1046 (*csn1*). SPy1044 is annotated in the NCBI database as hypothetical protein (NCBI, 2008). *Csn1* is encoding a CRISPR (clustered regularly interspaced short palindromic repeats) associated protein of unknown function (Fig.17 and 25, Table 10). For SpyRNA049 three different 5' ends were



determined corresponding to three transcripts originating from the locus with sizes of 171, 89 and 76 nt. The transcripts have the same 3' end sharing one rho-independent terminator.



**Fig. 9.** Primer extension analysis of SpyRNA049. Lanes A, C, G, T represent the sequencing reactions, lane PE OLEC703 represents the primer extension reaction done with primer OLEC703 and 8  $\mu$ g of TP2 RNA from EC904. 1:10 dilution of the radiolabelled primer was loaded alongside on the gel. The transcriptional starting signals are indicated by arrows on the right and the 5' start nucleotides are highlighted with a red frame. Only one 5' end was confirmed by self ligation (SL).

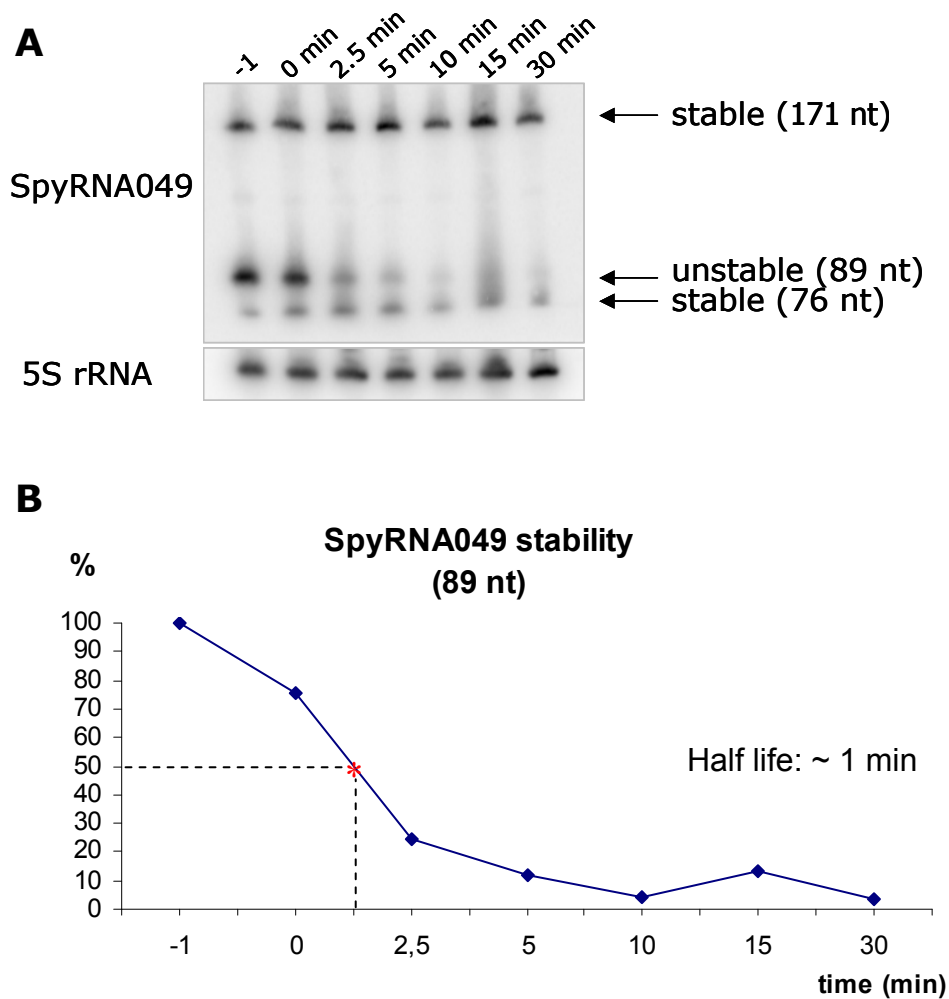


**Fig. 10.** Genomic localization of SpyRNA049 and its sequence. SpyRNA049 is encoded on the opposite strand of that from the adjacent genes; sequence is given from 5' to 3' of SpyRNA049 (minus strand). In the sequence, the putative promoter regions are highlighted in blue; 5' ends in red; putative rho-independent terminator in yellow.

## 6.4 Stability of sRNAs

To monitor the decay of the sRNA candidate, the activity of RNA polymerase was blocked by addition of rifampicin during cell growth, thus leading to the inhibition of *de novo* synthesis of RNA. Northern blot analysis was done using total RNA prepared from rifampicin-treated cells and the half life of the sRNA was determined by comparison to that of the stable 5S rRNA. 5S rRNA, like all ribosomal RNAs in a cell, is highly stable under normal growth conditions, and these conditions were therefore used for the determination of RNA stability (Deutscher, 2006). The time at which 50% of the RNA present in a cell was degraded is referred to as half-life of an RNA.

One example of the hybridization pattern is shown for candidate SpyRNA049 (Fig. 11A). A very interesting phenomenon is the difference in the stability of the three transcripts from the same locus. For SpyRNA049 (89 nt) a half-life of about 1 min was determined, however in contrast SpyRNA049 transcripts of 76 and 171 nt showed high stability with a half-life of more than 30 min.



**Fig. 11.** SpyRNA049 stability. (A) Northern blot hybridization of RNA from rifampicin-treated cells (10  $\mu$ g/lane) with SpyRNA049. (B) Half-life calculation of the unstable 89 nt transcript from the SpyRNA049 locus. Before rifampicin addition, the amount of the transcript is 100%. Over the time decreasing amounts of RNA are normalized with 5S rRNA amounts in the cells. The half-life is defined as the point at which 50% of the RNA were degraded (red asterisk).

Stability of other sRNA candidates was determined and results are pictured in Fig. 12. Riboswitches and all other RNAs encoded in 5' UTRs have a very short half life (< 30 sec), while functional sRNAs have a half life of more than 30 min. Interestingly, the 76 and 171 nt SpyRNA049 transcripts were as stable as the functional sRNAs.

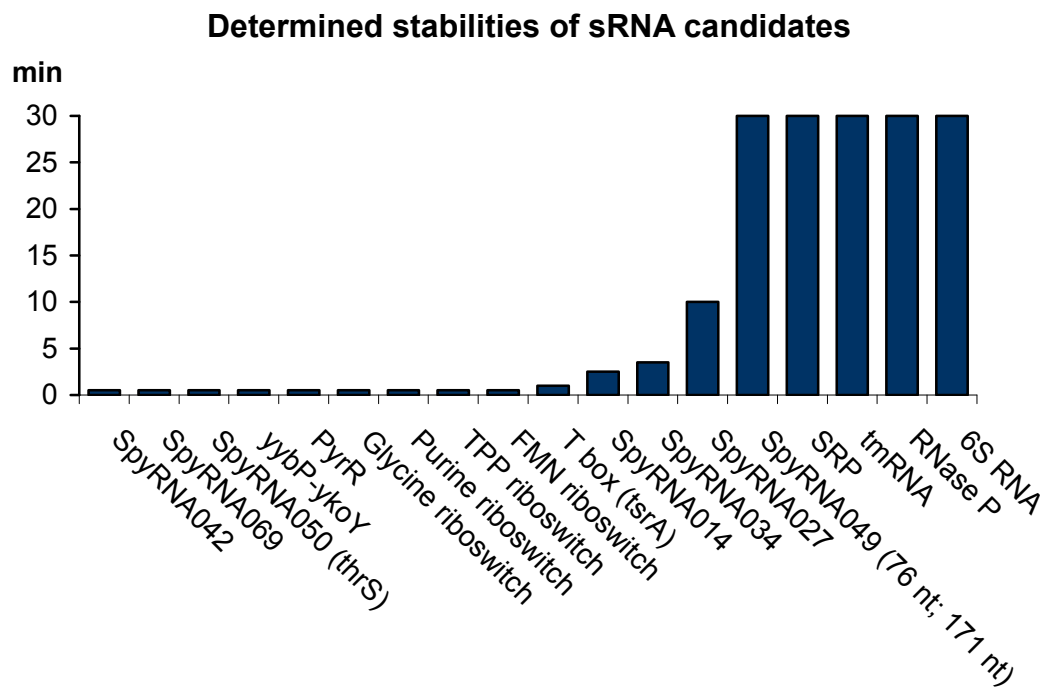
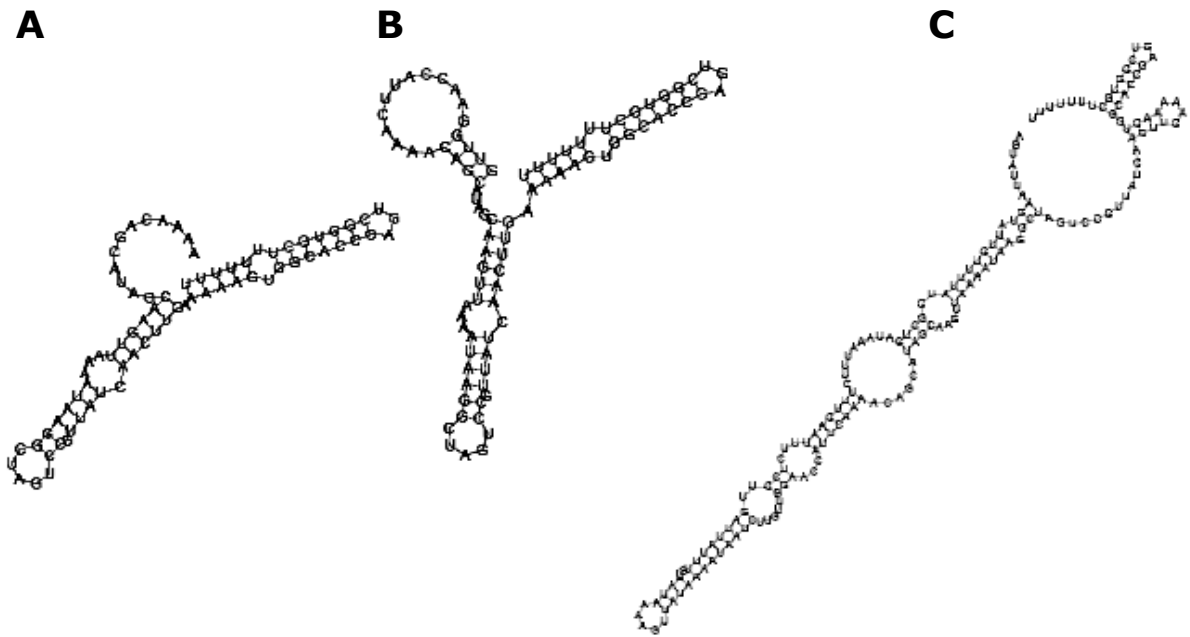


Fig. 12. Summary of results obtained from sRNA stability experiments using rifampicin.

## 6.5 Computational analysis of sRNAs

### 6.5.1 Structure predictions

Two-dimensional structure predictions of all sRNA candidates were done by Tanja Gesell and Maria Eckert using the computer programs Mfold and RNAfold (see 4.1.3). For SpyRNA049, structures for the three different sRNAs were predicted. The structures of the two shorter RNAs, 76 nt and 89 nt, are similar except for the 5' end where the longer RNA has an additional loop (Fig. 13). For the longest transcript (171 nt) a different structure was predicted that consists of a long hairpin loop with several bulges.



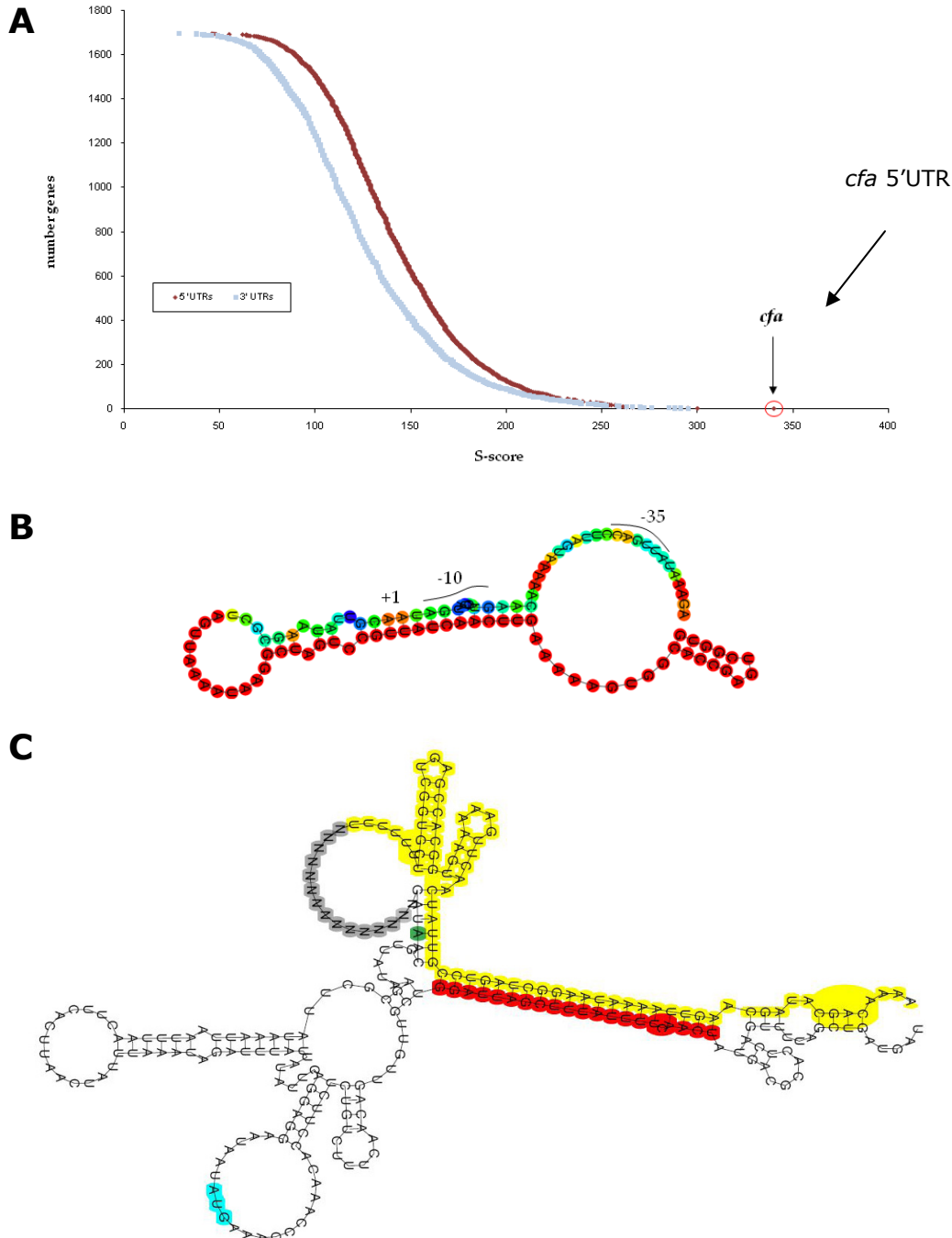
**Fig. 13.** Structure predictions for the three transcripts expressed from SpyRNA049 locus using the computer program RNAfold. (A) 76 nt (B) 89 nt and (C) 171 nt sRNAs. After RNA fold folding of SpyRNA049,  $\Delta G$  171 nt: -33.90 kcal/mol;  $\Delta G$  89 nt: -18.92 kcal/mol,  $\Delta G$  76 nt: -16.40 kcal/mol.

### 6.5.2 sRNA target predictions

A computer program (Mandin *et al.*, 2007) was used to predict putative mRNA targets of sRNAs. For this, 5' UTR and 3' UTR sequences of all annotated ORFs in the MGAS SF370 genome were analyzed for their potential interaction with the novel sRNAs. Predicted interactions between sRNA and mRNA with higher than average S-score were visualized using RNAcofold and RNAduplex (Vienna RNA Package). The prediction with the highest S-score for SpyRNA049 was the upstream region of the *cfa* gene. Unfortunately, the predicted interaction (Fig. 14A) is overlapping a putative promoter sequence (-10 region) of *cfa* (Fig. 14) (Gase *et al.*, 1999).

Hence, the interaction was verified by folding a hybrid RNA *cfa*-RNA049 with RNAfold (adapted from (Boisset *et al.*, 2007)), using the promoter of *cfa* (Gase *et al.*, 1999) and the three possible SpyRNA049 transcripts as starting point. Surprisingly, in all different hybrids tested, SpyRNA049 could bind *cfa* mRNA in the coding sequence, thirteen codons downstream of the ATG, thus forming a

long double stranded RNA stretch of sixteen nucleotides (Fig. 14C). Such an interaction was not expected and is raising many questions discussed below.

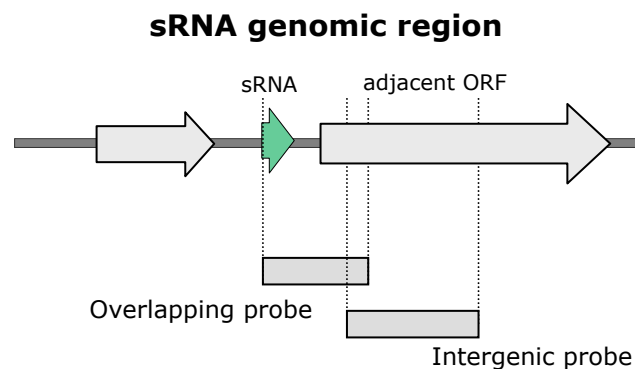


**Fig. 14.** mRNA target predictions for SpyRNA049. (A) S-scores for potential interaction between 3' UTRs (grey) and 5' UTRs (red) of the GAS genome with SpyRNA049. (B) RNAcofold predicted duplex structure of SpyRNA049 and the upstream region of the *cfa* gene. SpyRNA049: red. (C) RNAfold predicted interaction between full length SpyRNA49 (76 nt) and *cfa* transcript (150 nt). Linker: yellow; AUG: grey; interaction site: red.

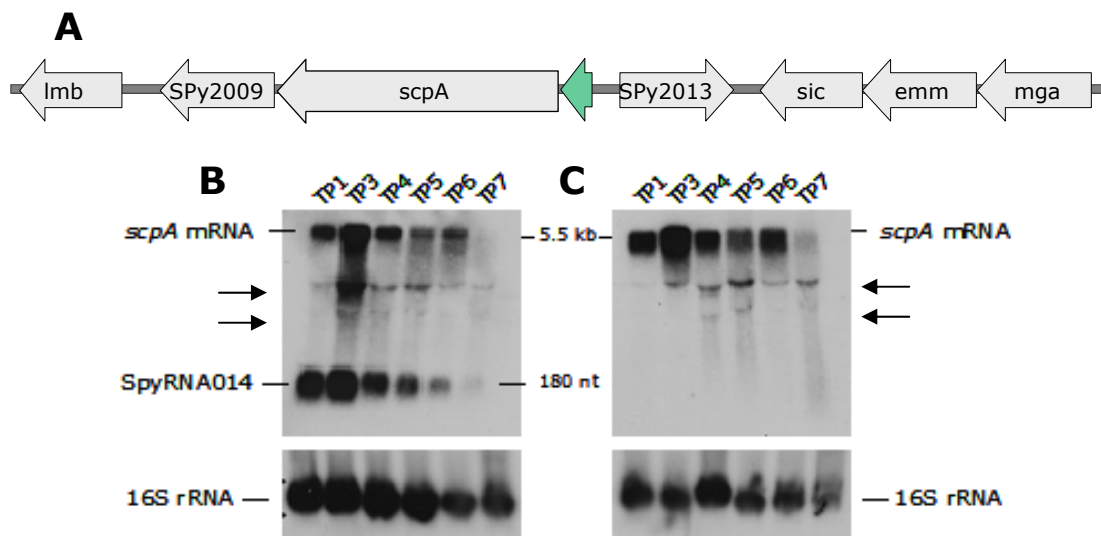
## 6.6 Further analysis of novel sRNA candidates

### 6.6.1 Comparative expression analysis of sRNA candidates with mRNAs of adjacent ORFs: the case of SpyRNA014

If a novel sRNA was encoded in the 5' UTR or 3' UTR of an mRNA (5 candidates), the expression of the adjacent gene was analyzed by Northern blots using two different probes. First, a probe specific to both the sRNA candidate and the downstream mRNA (overlapping sequence) was used. The second probe was detecting only the downstream mRNA (Fig. 15). As loading control for these experiments, probes detecting specifically the 16S rRNA or 5S rRNA were used. An example of a hybridization pattern is shown in Fig. 15 with the expression of candidate **SpyRNA014** and its downstream gene *scpA*.



**Fig. 15.** Probes used for the analysis of expression of the sRNA compared to that of the adjacent ORF.



**Fig. 16.** (A) Genomic location of *SpyRNA014* and surrounding genes. For exact size of ORFs please refer to Fig. 21. Northern blots, using agarose gels, hybridized with a radiolabelled DNA probe complementary to a sequence overlapping the predicted *SpyRNA014* RNA and the adjacent ORF *scpA* (B) or a sequence internal to the ORF of *scpA* (C). To each lane, 25  $\mu$ g of total RNA preparations from SF370 cells grown to 6 time points were applied. 16S rRNA was used as loading control. Bands that were due to unspecific hybridization are indicated by arrows.

*SpyRNA014* was of special interest, because it is located just upstream of the virulence gene *scpA*, encoding C5a-peptidase, with a known important role in virulence (please refer to 3.1.1.1). The growth phase dependent expression pattern of the *scpA* transcript and that of the small ncRNA *SpyRNA014* are similar (Fig. 16). Furthermore the mRNA transcript of *scpA* (5.5 kb) seems to correspond to a polycistronic messenger since the *scpA*-mRNA was expected to be approximately 3.5 kb in size.

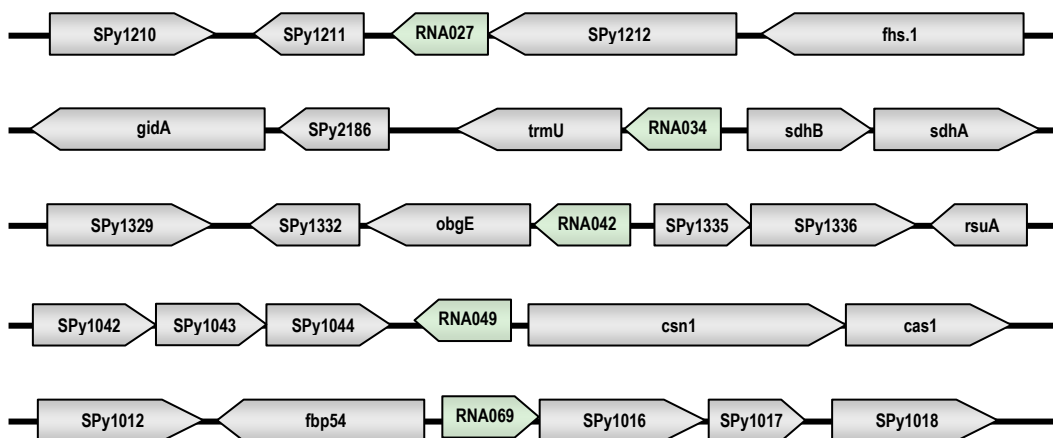
Unfortunately, for the other candidates, I could not highlight specific bands by Northern blot analysis due to a strong background (unspecific hybridization to the rRNA).



## 6.6.2 Genomic localization analysis of novel sRNA candidates

For all novel sRNA candidates, the flanking genes were further annotated with regard to their coding sequences, their orientation and their function (Table 10, Fig. 17; for detailed information refer to Fig. 21-26, Appendix).

In *S. pyogenes*, only three non-coding small RNAs have been described so far. In this study 5 new small non-coding RNAs were revealed.



**Fig. 17.** Genomic location of the 5 novel sRNAs identified with the bioinformatic screen. The sRNAs are shown in green and the adjacent genes in gray. The arrows indicate the orientations of the adjacent genes.

## 6.6.3 The case of SpyRNA049

### 6.6.3.1 Genomic localization and expression analysis of SpyRNA049

**SpyRNA049** is located in the 379 bp long intergenic region between the two ORFs SPy1044, annotated as hypothetical protein and *csn1*, a CRISPR-associated protein of unknown function. The CRISPR system is important in bacteria to acquire resistance against phage infection and consists of short direct repeats separated by non-repetitive spacers of similar size preceded by several genes encoding CRISPR-associated proteins (CAS proteins). It was suggested that the CRISPR system acts via an RNA interference (RNAi)-like mechanism. Phage sequence-derived spacers that would be processed off from a longer transcript using a special mechanism, would target phage

mRNAs or DNA for degradation with the help of CAS proteins (Sorek *et al.*, 2008).

As shown above, SpyRNA049 is transcribed independently and is not located in an UTR of an mRNA and thus could be considered as a typical non-protein-coding RNA. Further a stability of more than 30 min was shown for the 171 nt and the 76 nt long transcripts. In addition, using a computational approach to search for putative mRNA targets for the sRNA candidates led to the *cfa* mRNA, encoding the CAMP factor, as a possible target for SpyRNA049. For reasons above, we chose to study further the possible biological function of SpyRNA049.

**Table 10.** Novel small non-coding RNAs in *S. pyogenes*.

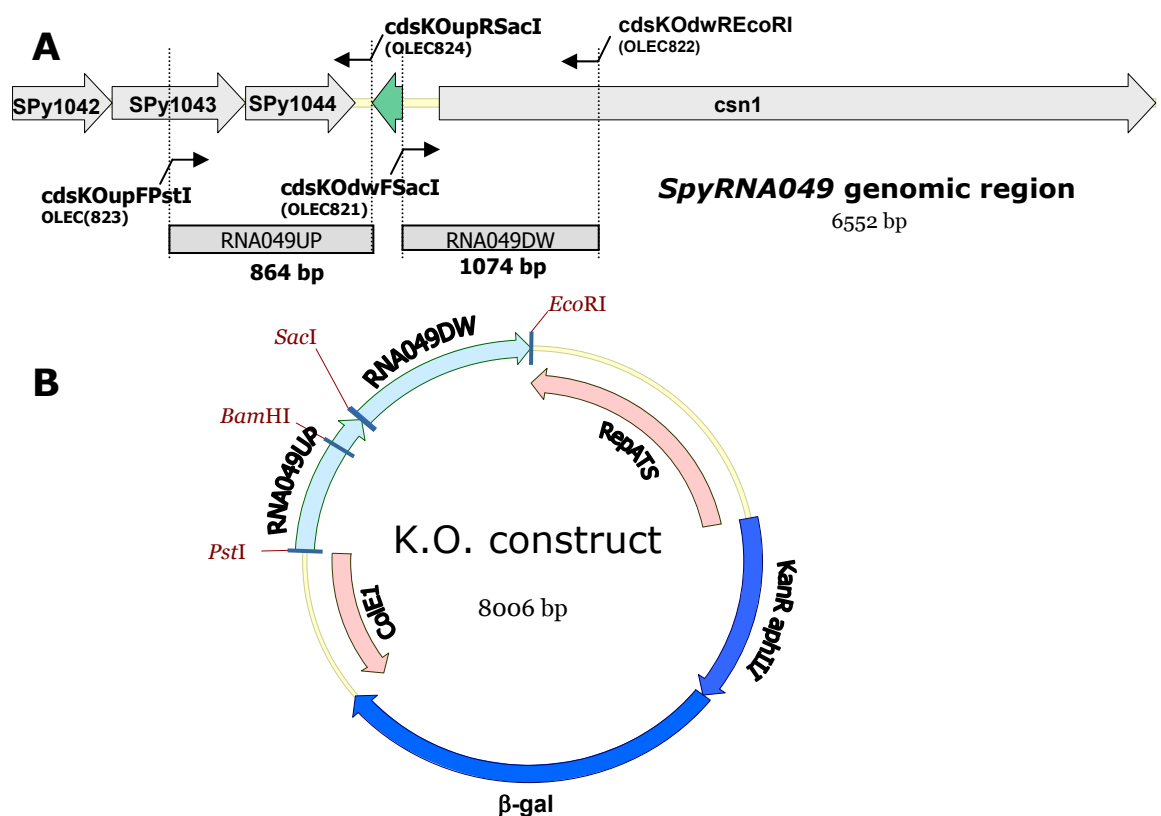
Candidate	Size of the sRNA candidate (nt)	Upstream ORF	Downstream ORF
SpyRNA027	130	<i>SPy1211</i> : 357 nt hypothetical protein	<i>SPy1212</i> : 1578 nt put. cardiolipin synthetase
SpyRNA034	110	<i>trmU</i> : 1122 nt tRNA-methyltransferase	<i>SdhB</i> : 672 nt put. L-serine dehydratase
SpyRNA042	195	<i>obgE</i> : 1314 nt GTPase ObgE	<i>SPy1335</i> : 516 nt conserved HP - transposon
SpyRNA049	80, 90, 170	<i>Spy1044</i> : 633 nt hypothetical protein	<i>csn1</i> : 4107 nt CRISPR-associated protein Csn1
SpyRNA069	190	<i>fbp54</i> : 1653 nt putative fibronectin-binding protein-like protein	<i>Spy1016</i> : 999 nt hypothetical protein

### 6.6.3.2 Creation of a SpyRNA049 mutant

To assess a possible role for SpyRNA049, I started to create a SpyRNA049 mutant strain in the M1 serotype strain EC904. The strategy to obtain the mutant was to delete the DNA region on the chromosome in-frame starting from the transcriptional start to the rho-independent terminator of the sRNA. The regions upstream (RNA049up) and downstream (RNA049dw) of SpyRNA049 (Fig. 18A) were amplified by PCR using Phusion polymerase. The RNA049up fragment was cloned, using PstI and SacI restriction enzymes, into the pEC214 vector, a *S. pyogenes* - *E. coli* shuttle vector containing a thermo-sensitive origin of replication

(*S. pyogenes*), a non-thermosensitive origin of replication (*E. coli*) and a kanamycin resistance cassette (selection in *E. coli* and *S. pyogenes*) thus generating plasmid pEC262 (Table 3). As a next step it was attempted to clone the RNA049dw fragment, using *SacI* and *EcoRI* restriction enzymes, in pEC262 (Fig. 18A) and then to introduce the plasmid in *S. pyogenes* by electroporation.

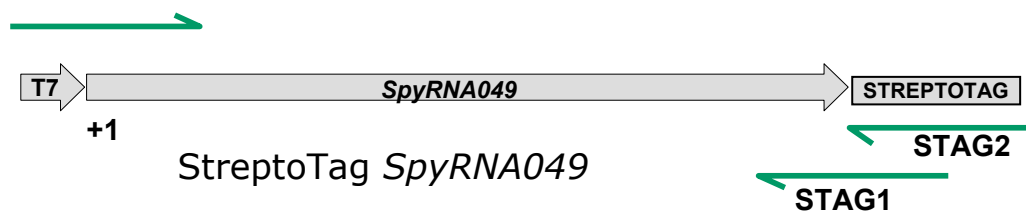
Since the vector is temperature-sensitive, the plasmid can only replicate at 28°C. At 37°C, the plasmid cannot replicate and gets lost within the following cell divisions or integrates into the chromosome under antibiotic selection. The excision of the plasmid from the chromosome is stimulated by growth at permissive temperature (28°C) in the absence of antibiotic. This leads to the generation of a mixed population of wild-type and knock-out clones. Knock-out clones are selected further using PCR analysis and analyzed by Southern blot, PCR and sequencing analysis for verification.



**Fig. 18.** (A) Genomic location of *SpyRNA049* indicating the upstream and downstream fragments used to construct the *SpyRNA049* deficient vector (grey boxes), the primers (arrows) and the surrounding genes. (B) Recombinant shuttle plasmid with pEC214 vector backbone containing the upstream and downstream fragments of *SpyRNA049*.

### 6.6.3.4 Fishing proteins binding SpyRNA049 using the StreptoTag method

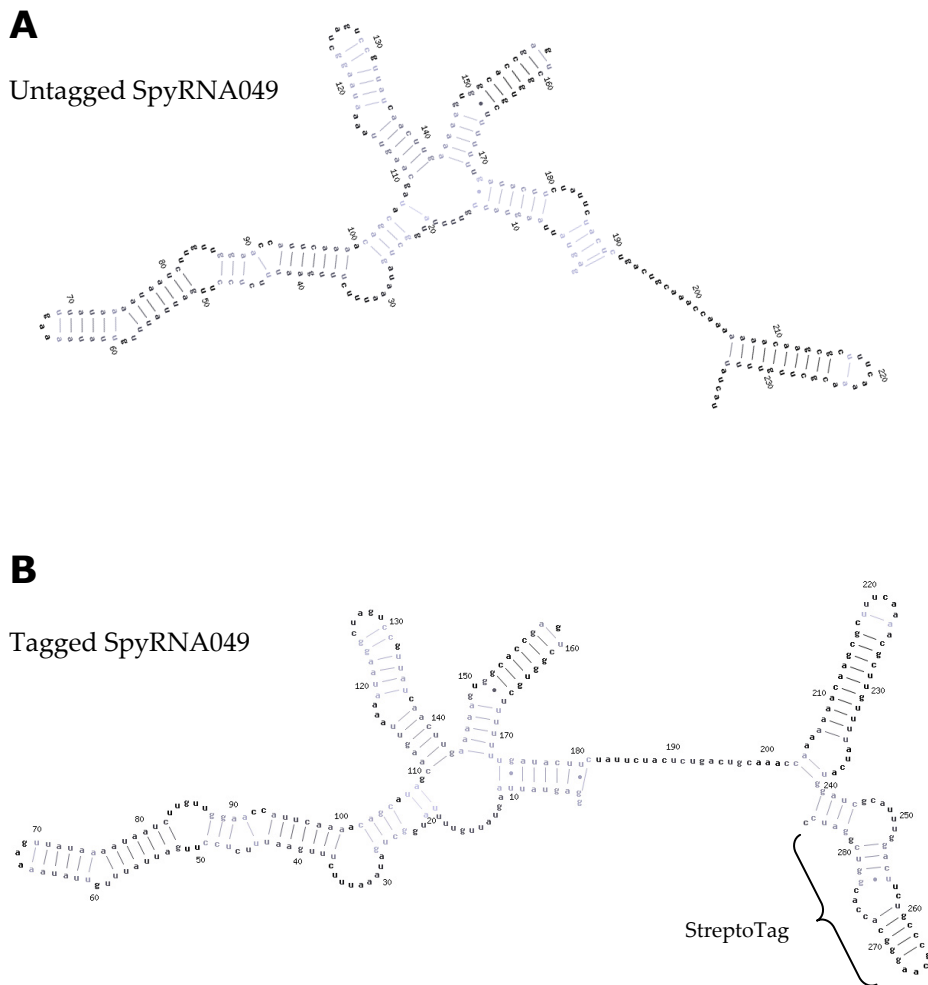
Since so far no proteins in *S. pyogenes* are known that specifically bind sRNAs, like a Hfq homologue, the set up of a new method (StreptoTag), was started. StreptoTag is a one-step affinity purification method to identify RNA binding proteins with high and low binding affinity (Bachler *et al.*, 1999). First, a streptomycin aptamer-sRNA hybrid needs to be created. For this, a short streptomycin aptamer is fused to the 3' end of an sRNA of interest by using a two-step PCR method. The generated amplified DNA fragment is then used as template for *in vitro* transcription (Fig. 19). We wanted to apply this technique to fish potential proteins binding SpyRNA049. A forward primer containing a T7 RNA polymerase promoter sequence and two reverse primers were used to add the StreptoTag to the RNA molecule (Fig. 19). One PCR was performed with the forward primer and the first reverse primer (STAG1) using EC904 genomic DNA as template, then the second PCR was done with the same forward primer and the second reverse primer (STAG2) using the PCR product of the first PCR as template. A tagged 6S RNA was also prepared. 6S RNA was used as control, since it was described in other bacteria as a protein binding RNA (Wassarman, 2007) and its binding partner (RNA polymerase) is a binding protein candidate that should be easily isolated and purified.



**Fig. 19.** PCR construct used for *in vitro* transcription of tagged SpyRNA049. The primers used for the two step PCR are indicated (green arrows).

To check if the predicted secondary structures of the RNA molecules were not affected by the addition of the tag, the program Contrafold

(<http://contra.stanford.edu/contrafold/>) was used. For SpyRNA049 (Fig. 20) and 6S RNA, no secondary structure disturbance was found. Each hybrid RNA molecule was immobilized onto streptomycin coupled sepharose columns and then total cell extracts of *S. pyogenes* EC904 grown to early-mid-log phase (TP 2) were passed through the column, allowing proteins to bind the RNA. Since so far no phenotype for SpyRNA049 is known, for protein preparation a timepoint was chosen where the RNA is highly expressed (TP2). The protein lysates were always prepared freshly before use. Crude *S. pyogenes* extract containing proteins were successfully prepared and checked by SDS-PAGE. For all prepared lysates results were obtained between OD<sub>595 nm</sub> of 0.49 and 0.53, corresponding to a protein concentration of 1.63 to 1.77 µg/µl. After several washing steps the hybrid RNA with bound proteins was eluted from the column with streptomycin. Proteins isolated using this method can be analyzed by SDS PAGE or by mass spectrometry (Windbichler and Schroeder, 2006). The set up of the method could not be finished in time to obtain meaningful results.



**Fig. 20.** Secondary structure predictions of untagged (A) and tagged (B) SpyRNA049 (Contrafold).

---

## 7. Discussion

---

The importance of small non-coding RNAs in regulatory processes became recently an important topic in molecular biology. It is very likely that these small, non-coding RNAs play an important role in regulating virulence in many bacterial pathogens, such as the already known RNAIII in *S. aureus*, which controls globally the expression of several virulence factors (Huntzinger *et al.*, 2005).

### 7.1 Computational approach to search for novel sRNAs in *S. pyogenes* and experimental validation using Northern blot screen

To evaluate the existence of novel small non-coding RNAs in *S. pyogenes*, a wide scale computational search was done. Three different bioinformatical programs (QRNA, RNAz and Alifolds) were used by bioinformaticians of two different institutes of the University of Vienna to predict intergenic regions in the genome of *S. pyogenes*, in which potentially sRNAs could be encoded. These predictions were compared to already described predictions (sRNAPredict2 (Livny *et al.*, 2006) and Rfam database (Griffiths-Jones *et al.*, 2005)) and were subsequently experimentally validated. In total the predictions resulted in 90 putative RNA loci (Table 12) that were experimentally tested with Northern blot analysis and 29 of them gave a positive result (Table 8).

These 29 positive candidates are 5 novel sRNAs, which were not described previously, SpyRNA014, which was already once mentioned in the literature, 5 leader RNAs, 6 riboswitches, 4 functional RNAs (housekeeping RNAs) and 8 T boxes. Of the 5 novel sRNAs, 3 were encoded in 5' UTRs of the downstream mRNAs (SpyRNA034, SpyRNA042 and SpyRNA069), one was encoded in a 3'

UTR (SpyRNA027) and one candidate was encoded independently in an IGR (SpyRNA049).

It is important to keep in mind that these computer programs are not 100% reliable for small RNA predictions and it should not be assumed that in this study all existing sRNAs in *S. pyogenes* would be identified. One reason for this is that the predictions are limited to regions of the genome, which were conserved throughout evolution. Unfortunately none of the programs used was able to predict one of the already known sRNAs in *S. pyogenes*: *fasX*, *rivX* or *pel* RNA. These programs also predicted sRNA candidates, the expression of which could not be detected by Northern blot analysis with our experimental settings in this study. In this context, it should be mentioned that Northern blot analysis used in this study is not a very sensitive method to detect the expression of RNA. Therefore, only sRNAs that are expressed in abundance under normal growth conditions were identified. One can not rule out the possibility that the sRNAs that were not shown to be expressed using Northern blot analysis, could be nevertheless expressed in *S. pyogenes* under certain conditions of growth. Another possibility to test the expression of sRNAs would have been the use of reverse transcription (RT) PCR, but a Northern blot screen was preferred because the expression profile and the approximate sizes of the RNAs can be determined in one single experiment. In the Northern blot pictures some RNAs showed a hybridization pattern of more than one band. Some of the multiple bands could be correlated to several 5' ends using primer extension analysis as it was the case for SpyRNA049. In general, multiple bands detected on Northern blots can also be assigned to RNA fragments resulting from maturation or degradation mechanisms, which involve processing of the RNA.

## **7.2 Characterization of expressed sRNA candidates**

The expressed sRNA candidates were further characterized by determining their 5' and 3' ends, their metabolic stability, their predicted secondary structure and the prediction of putative mRNA targets. Expression of the sRNAs was also



analyzed with regard to the expression of genes located upstream and downstream of the sRNA.

First, the exact genomic location of the sRNAs was determined by mapping their 5' and 3' ends. For most of the candidates only one +1 starting signal was found. An exception was for SpyRNA049, which showed three clear signals for three different 5' ends. These starting signals correlated with the three different sRNAs that were revealed by Northern blot. In the literature, the sRNA *rivX* from *S. pyogenes* shows as well three starting signals but *rivX* RNA is cotranscribed with *rivR* and since no consensus promoter sequences were found directly upstream of *rivX* RNA it was speculated that *rivX* RNA could be most likely processed off the transcript (Roberts and Scott, 2007). Since both genes adjacent to SpyRNA049 are encoded on the opposite strand, cotranscription with another gene seems to be quite unlikely. Surprisingly, there is one putative promoter consensus sequence (-35 and -10) found upstream of the +1 signal that would correspond to the 89 nt long sRNA (Fig. 10). Only a second (-10) putative promoter sequence was identified upstream of the +1 signal that would correspond to the 171 nt long sRNA.

Following the mapping of the sRNAs, the stability was determined in rifampicin experiments. In this study it was clearly shown that sRNAs encoded in untranslated regions of mRNAs have a very low stability in the cell that ranges from about 30 seconds to a few minutes. In contrast, housekeeping RNAs that were predicted by the Rfam database like 6S RNA, tmRNA or SRP are highly stable and also known regulatory RNAs in *S. pyogenes* like *fasX* and *pel* RNA have clearly higher stabilities than sRNAs encoded in UTRs of mRNAs. The candidate of this study with the highest stability of all the novel sRNAs was candidate SpyRNA049, where the longest RNA of 171 nt has a stability of more than 30 min (Fig. 12). This might indicate that this RNA has a regulatory function in the cell. Because of these results and the special location of SpyRNA049 as the only novel sRNA of unknown function not encoded in an untranslated region of an mRNA, this sRNA was chosen for further investigation.

### 7.3 Analysis of SpyRNA049

The computer program described by Mandin *et al.* (2007), which analyzes 5' and 3' UTRs of mRNAs as potential interaction partners for an sRNA of interest was used to predict mRNA targets of six sRNA candidates (SpyRNA014, SpyRNA027, SpyRNA042, SpyRNA049 and SpyRNA069). For SpyRNA049 an interesting result was obtained, since this program revealed the 5' UTR of the *cfa* mRNA as potential target. The *cfa* gene encodes the CAMP factor, a protein known to participate in the CAMP reaction that was used for the classification of bacteria. In the CAMP reaction, which was first described by Christie *et al.* (1944), erythrocytes are lysed by the synergistic action of the CAMP factor secreted by Streptococci and the  $\beta$ -toxin of *Staphylococcus aureus*, which is a sphingomyelinase (Gase *et al.*, 1999). The term CAMP originates from the first observers of the phenomenon, Christie, Atkins and Munch-Petersen (Brown *et al.*, 1974). The CAMP factor was more extensively studied in group B streptococci (GBS) and is known to oligomerize and form pores in susceptible membranes. It also showed toxic effects when injected intravenously in mice and rabbits (Skalka and Smola, 1981). Data, which suggested that CAMP factor would bind to Fc fragments of immunoglobulins could not be confirmed in a later study (El-Huneidi *et al.*, 2007). Recently a GBS  $\Delta cfb$  mutant strain was generated and in this study, no effect on systemic virulence could be observed in terms of resistance to phagocytic cells, endothelial cell invasiveness and virulence in a mouse infection model. Still, the role of CAMP factor in physiological aspects and in virulence needs to be elucidated. In *S. pyogenes* so far only the gene *cfa* itself was described (Gase *et al.*, 1999). The *cfa* gene of GAS SF370 has 67% identity to *S. agalactiae cfb* and 65% to the *cfu* gene of *Streptococcus uberis*. It was shown that expression of *cfa* is under the control of a strong promoter that might be regulated at the transcriptional level by a so far unknown factor (Gase *et al.*, 1999). Candidate SpyRNA049 was predicted to bind to an upstream region of the *cfa* ORF overlapping the predicted promoter region of *cfa* mRNA. Unfortunately, the exact 5' end of the *cfa* mRNA was so far not experimentally determined. This

prediction can not be explained by a so far known mechanism of sRNAs. One possibility would be that the *cfb* gene is also transcribed from another predicted promoter lying further upstream, which would suggest an RNA-RNA interaction. In this case a direct Shine-Dalgarno sequence sequestration seems very unlikely regarding the prediction of the binding site of SpyRNA049 to the *cfb* mRNA with the RBS lying further downstream. On the other hand, the RNA duplex of SpyRNA049 and *cfb* mRNA might serve as target for RNases, which leads to the degradation of the complex. Another possibility might be that SpyRNA049 is binding directly to the DNA in the predicted region, although so far no direct sRNA-DNA interaction *in vivo* has been shown. A direct interaction with DNA might be responsible for a regulation at the transcriptional level. Because the transcription of a *cfb* mRNA is supposed to be initiated at the proximal promoter, we investigated further the capability of SpyRNA049 to bind *cfb* mRNA. Surprisingly, SpyRNA049 was predicted to bind *cfb* mRNA forming a long double stranded (ds)RNA stretch. We did not expect such a strong binding. However, even though the dsRNA stretch seems to be located too far away from the RBS or the ATG of the *cfb* mRNA to hinder ribosome binding, it might impede polysome formation on the messenger, then leading to decrease translation efficiency. On the other hand, this interaction could as well trigger a RNase III-mediated degradation having as consequence a depletion in the mRNA pool and thus leading to decreased protein synthesis. To analyze whether the sRNA SpyRNA049 interacts with the *cfb* mRNA, gel shift assays or footprinting analysis could be done.

To determine a possible biological function of SpyRNA049, the first steps towards the creation of a mutant strain were done. The strategy used was to delete the SpyRNA049-encoding DNA region on the chromosome from the first transcriptional start to the rho-independent terminator at the 3' end, without the insertion of an antibiotic resistance cassette. For this procedure, a vector is needed containing the upstream and downstream region of the sRNA sequence that should be deleted from the chromosome. This vector is first integrated and

then excised from the chromosome by homologous recombination events induced by temperature shifts. Unfortunately, the cloning steps did not function as expected such that only the DNA fragment upstream of the SpyRNA049, designed for the creation of the deficient mutant, was inserted in the thermosensitive shuttle vector. The next steps would be finishing the mutant, verifying it (verification by PCR, Southern blot and sequencing analysis) and the study of a potential phenotype. It still needs to be verified if SpyRNA049 is a real 'non-protein-coding' sRNA since there is one 69 nt long putative ORF (23 amino acids) found in the region of this RNA. However it seems unlikely that a small peptide is encoded because no RBS was predicted upstream of this putative start codon of this putative peptide.

## 7.4 Additional analysis of SpyRNA014

The transcriptional expression profile of the genes adjacent to the novel sRNAs was examined by Northern blot analysis (agarose gels, Table 9). As an example, SpyRNA014 is encoded in the 5' UTR of the *scpA* gene and has already been mentioned in the literature as a truncated *scpA* mRNA, generated by a transcription termination mechanism (Pritchard and Cleary, 1996). SpyRNA014 was still included in this study since the mechanism of *scpA* regulation was so far not elucidated and other functions of SpyRNA014 cannot be ruled out.

## 7.5 Outlook

Taken together, the large scale computational approach employed in this study allowed the identification of 29 sRNAs in *S. pyogenes*. The next step is the determination of the functions of the identified sRNAs and the characterization of their mode of action, which is a great challenge. In terms of SpyRNA049 the first steps in this direction have been made through the creation of an isogenic deficient mutant. On the other hand, the sRNAs could also be overexpressed, for description of a phenotype. Further analyses are necessary to proof that the

positive candidates are in fact non-protein-coding sRNAs with regulatory functions, excluding a protein coding function of the putative ncRNA. Especially in *S. pyogenes*, which is one of the most frequent pathogens, knowledge about sRNAs is very limited including their targets and binding partners, which all might serve as potential drug targets. The topic of small ncRNAs has recently emerged and there is definitely still a lot to discover.

## 8. Appendix A

**Table 11.** Table of primers used in this study.

sRNA candidate	Oligo	*F/ R	Sequence 5' - 3'	Expected PCR product	T* (°C)	Used as
SpyRNA001	OLEC342	F	GTGTTATCTGTGGGATTACA	214 bp	55	probe
	OLEC343	R	TCAAGTGATGGTGTTC AAC			
	OLEC344	F	GAAAAAGCCCTTCTTGAA			
SpyRNA002	OLEC345	R	ACAGCAATGACTGGAGATAA	224 bp	55	probe
	OLEC346	F	AGTAATAACTATTGTGCGTCC			
	OLEC347	R	CAAACCCTATCTTTCTAAA			
SpyRNA003	OLEC348	F	ACTCCTCTCACAATAACAGA	129 bp	55	probe
	OLEC349	R	GTTATGCTTTGCTTGTTAA			
SpyRNA004	OLEC350	F	CGTATAGAACGCAAAGTGAA	168 bp	55	probe
	OLEC351	R	TTAGCAAAAACGGATAGGCA			
	OLEC371	F	TCTTTTGTGTCAGTGGCTTGTA			
	OLEC372	R	TCATTGAACTCAGGCTAAAC			
	OLEC373	F	TCITTACCAACAAATATTTA			
SpyRNA005/yypP-ykoY	OLEC374	R	GTTTACTCTTTCATTATAGC	81 bp	55	probe
	OLEC352	F	AGATAGCAGTAAATCATCTGA			
	OLEC353	R	GTTGTTCAAATAAAAAAGACC			
SpyRNA006	OLEC377	F	AGAAGTATTTATGATTCGGA	195 bp	55	probe
	OLEC378	R	GACCAAACTATCATTTCAGA			
SpyRNA007	OLEC379	F	ATTGTGGAATAGTTCTGAGA	320 bp	55	probe
	OLEC380	R	ATTATCATTGACAGGTGAGC			
SpyRNA008	OLEC381	F	TTTCTAGGTTAGTCATGGGA	408 bp	55	probe
	OLEC382	R	TTAACACGTCCCGTAACAAA			
SpyRNA009	OLEC383	F	CACTCTTTGTGGAGATCAG	243 bp	55	probe
	OLEC384	R	GCTAGATAAAGGGATAGCCG			
SpyRNA010	OLEC385	F	GGTGGGACTAAGGTTACCTT	473 bp	55	probe

	OLEC386	R	AAGTACTCAGCAATCAAAGC			
	OLEC387	F	TGATTGCTGAGTACTTCACA		55	probe
	OLEC388	R	ACTGGGTCAGCTTTATAAGT	346 bp		
	OLEC389	F	ATAAAGCTGACCCAGTTTCA		55	probe
	OLEC390	R	CGTCATGTGATCGTTAGTAA	404 bp		
	OLEC391	F	CCGCCAACCAACAGAACAAG		55	probe
	OLEC392	R	AATGTGGTTTTAATTAGACC	395 bp		
SpyRNA011/TPP	OLEC397	F	CCATAAGAAGGACTAAACCA		55	probe
	OLEC398	R	TAAAACCCCTTATCCTTGAC	311 bp		
	OLEC447	F	AAGAGTATCGCTTTCCA	with OLEC398: 175 bp	55	probe
SpyRNA012/PyrR1	OLEC399	F	AGCCAGGAAATATGAGAAGTC		55	probe
	OLEC400	R	GCGCGTTTCATGGTCACG	352 bp		
SpyRNA013	OLEC401	F	ACAGGTGCATCAGTTAAGAA		55	probe
	OLEC402	R	TTGCAAGTGAAAATTCGCGA	472 bp		
SpyRNA014	OLEC403	F	TTTTTACGCAATGTGTCGTC		55	probe
	OLEC404	R	TAACATAGGAGAACCTCACA	450 bp		probe
	OLEC513	R	CATTTAATCATTATGACAAAA	with OLEC403:199 bp	55	probe
	OLEC514	F	TATCTCATTTATCAATTA		55	probe
	OLEC515	R	GTTCTCTCTTTTATGTCTA	134 bp		
	OLEC516	F	CGATCGGTTCTCTCTTCA	with OLEC404:117 bp	55	probe
	OLEC556	F	TTTGTCTTTATGACCATTAAGAAAAAGAGAG		55	Seq./PE**
	OLEC555	R	GGAGGTCACAAACTAAGCA	260 bp		
	OLEC642	F	TGTCTAAAAGAATGAGGATAAGGAGGTCAC		55	Seq./PE**
	OLEC643	R	CAGGGGTTATTTGCATATGATACAGCTA	180 bp		
SpyRNA015	OLEC405	F	GATATACCAGCCTTTGCGAC		55	probe
	OLEC406	R	GCTCTCTATAGCATTAAAGGA	334 bp		
SpyRNA016/L20 leader	OLEC407	F	AGCTCGTGATTACAAGGATA		55	probe
	OLEC408	R	TCAGACAATGGTTTAATACC	419 bp		
SpyRNA017	OLEC409	F	TTCAGGCCAGTCATCTCCAG		55	probe
	OLEC410	R	ACTGAGTGAACATCAAGAGA	334 bp		
	OLEC494	F	TATTCTCTCCTTAATAAATTA	with OLEC410:168 bp	55	probe
	OLEC541	F	GTATTCTCTCCTTAATAAATTA		55	probe
	OLEC542	R	GCTATCACTTTGTAATACTGA	186 bp		
SpyRNA018	OLEC411	F	TATAGTTGCATTGTAATGGG	382 bp		probe

SpyRNA019	OLEC412	R	CGTCTACCGCCTTTTATATA			
	OLEC413	F	CAAAAAAAAAACACCACAGTAGA		55	probe
	OLEC414	R	TTGTTTAAACCAGGTCTAAA	364 bp		
SpyRNA020	OLEC495	R	ATTGAAATTGATTGAGATAAC	with OLEC413:447 bp	55	probe
	OLEC415	F	TGATACAACACCACCAGTTAC		55	probe
	OLEC416	R	AAGAAGATGAAGAACCAGTTC	321 bp		
	OLEC582	F	AAGCATTGCTTTCTTGTTA		55	probe
SpyRNA021/ileS T box	OLEC583	R	ATTATCATTTGACAAATTGTC	203 bp		
	OLEC417	F	ACCTGCGCGCATTGGAAAGGC		55	probe
	OLEC418	R	TCTCAATCTGATTCTACGAC	401 bp		
SpyRNA022/L19 leader	OLEC419	F	TAGCATATTTGTACCTCG		55	probe
	OLEC420	R	GGACGGAAGTTAGGGATA	307 bp		
	OLEC505	F	AGAAGCTTTTTGTTAGAATAA		55	probe
SpyRNA023	OLEC506	R	CTCCTATCTTACTAATCTTA	96 bp		
	OLEC421	F	TATTGCAACCTATGGTTTGA		55	probe
	OLEC422	R	GACTTAGAACAAGCGATTAC	330 bp		
SpyRNA024	OLEC423	F	AGAGAACTCCGTTTGAAAGA		55	probe
	OLEC424	R	AGTCCTCCTCTTTAGAATTC	401 bp		
SpyRNA025	OLEC425	F	ATGATGGTACGCTGTTTATA		55	probe
	OLEC426	R	TACCTCTAGAGTATGTTGAC	310 bp		
SpyRNA026	OLEC427	F	TGAACGGTTTGAGTATTTGA		55	probe
	OLEC428	R	CGCCATAATGTAAACTACAC	460 bp		
	OLEC496	F	TAGAAAAGAAGGGCTCTTGC	with OLEC428:380 bp	55	Seq./PE**
SpyRNA027	OLEC429	F	ATCAGCAACGTGGCCATCAC		55	probe
	OLEC430	R	TTAGTACGAATAATTGCCCC	405 bp		
	OLEC511	F	TCATCTCTACCTACATTATGC		55	probe
SpyRNA028	OLEC512	R	ATTGTGACTGCAAAGTCA	274 bp		
	OLEC661	R	GGGCATTCTTAATAAATAAGA		55	PE**
	OLEC439	F	TGGAAAGGCATAAACTGATA		55	probe
	OLEC440	R	CTACTCCTTTTCGATAAGAC	388 bp		
SpyRNA029	OLEC441	F	ATACTTTCGTTTCCCTGAC		55	probe
	OLEC442	R	CTTGAGTTTCTCAAGACGACG	446 bp		
SpyRNA030	OLEC443	F	AAGGCATAGAGTAACTTGGA	444 bp	55	probe



SpyRNA031	OLEC444	R	TCCTAGGGATCCTGCTAGTA	154 bp	55	probe
	OLEC450	F	GTTCATGAGTTCCGTTATC			
SpyRNA032	OLEC451	R	GCTATGGGGAATTTAGTAC	152 bp	55	probe
	OLEC452	F	GCGCTTTATATTCTGAAAAC			
SpyRNA033	OLEC453	R	AGATACGACAGTCATTATGC	411 bp	55	probe
	OLEC454	F	TTCCTGTTGACCTTGCCTTA			
SpyRNA034	OLEC455	R	TGGTAAAGGCTGTCTACTCC	172 bp	55	probe
	OLEC456	F	TAAGATAGAAAAAGATAATCA			
SpyRNA035	OLEC457	R	ATAATAAAAAATTATTGAGCA	386 bp	55	probe
	OLEC458	F	GTGATTCTAAAATCCAGTCC			
SpyRNA036	OLEC459	R	GGCGTTCCTACTTGTAGAC	379 bp	55	probe
	OLEC460	F	CCAGCACCTTAATGTACTGA			
SpyRNA037	OLEC461	R	CCTTTTACTGGACTCATCTA	182 bp	55	probe
	OLEC462	F	GACCTCACTTACTTTCCAAAC			
SpyRNA038	OLEC463	R	AATGAAAACGTTAGAAATTCA	342 bp	55	probe
	OLEC464	F	TGTGAAGGCTTCAGAAAGTA			
SpyRNA039	OLEC465	R	TGAGTGTAGTGATATGGGAA	390 bp	55	probe
	OLEC466	F	GATCAAGGAACAGTCAGATC			
SpyRNA040/trsA T box	OLEC467	R	TCCCTGTGTCAGTGTCAATC	333 bp	55	probe
	OLEC468	F	AAAAAGTCCCCACGCTATTA			
SpyRNA041	OLEC469	R	TGTA CTGAGACCGTCAAGTC	277 bp	55	probe
	OLEC470	F	GTGCTCAGAACGCAGACAAA			
SpyRNA042	OLEC471	R	CGGAGATTTCCCTCGCTACT	381 bp	55	probe
	OLEC472	F	CGCCTTTCAGCGTTCCCTTAC			
SpyRNA043	OLEC473	R	AATAGCACCCCAATCGTTAGA	189 bp	55	probe
	OLEC474	F	TGCATTTTCTCTTATCCA			
SpyRNA044	OLEC475	R	TTTACTAAGCCTCGTTGTC	297 bp	55	probe
	OLEC476	F	CGTCTTTGTCTTATTAAC			
SpyRNA045	OLEC477	R	AGGTAGAACGAGTCAATTTA	119 bp	55	probe
	OLEC478	F	AGGTATCAGGTGCTCCCTGAG			
SpyRNA046/valS T box	OLEC479	R	TAAAAAATCCTCCAACAAAAG	313 bp	55	probe
	OLEC480	F	CCTTGTCCCCTTTTCGTTGA			
SpyRNA047	OLEC481	R	GCATTGAGTTGGTCTGTTAC	382 bp	55	probe
	OLEC482	F	AGATATATTA AAAATATTACA			

	OLEC483	R	TAAAAGATAGTTTAGCTATTG			
SpyRNA048	OLEC484	F	GAGAGCGAAATCCCGATTTA		55	probe
	OLEC485	R	AACAGCCAAGGAGTACTGTC	332 bp		
SpyRNA049	OLEC486	F	TAAAACAAGCGTTTTGAAAAGC		55	probe
	OLEC487	R	TTAATGACCTCCGAAATTAGT	377 bp		
	OLEC703	R	CGACTCGGTGCCACTTTTTCA		55	PE**
SpyRNA050/thrS T box	OLEC488	F	TGAAGACAAGTTAGATTGGA			probe
	OLEC489	R	AGTGAATAGCTCAATACCAC	312 bp		
	OLEC624	R	ATAAGGGGGCCACCCTTTTATGTTTGCATA		55	Seq./PE**
	OLEC625	F	TCACCTCAGTAAAAGTAGTCA	228 bp		
SpyRNA051	OLEC490	F	CGGACACCATCAAATAGCA		55	probe
	OLEC491	R	TTACTCCCGCTGTAGTGACA	269 bp		
SpyRNA052	OLEC492	F	AGAGAATCGTTTAAACCCTGC		55	probe
	OLEC493	R	CAGCAATGTAGCTCTGATA	236 bp		
SpyRNA053	OLEC517	F	CCTCTTCAATTAAGTCTGA		55	probe
	OLEC518	R	ACTACTAGCATGTCCTAACC	225 bp		
SpyRNA054	OLEC519	F	GGCTGTCCTTTCGTTTTTGTGTC		55	probe
	OLEC520	R	CGATTTAAATAAGCCCCATCA	241 bp		
SpyRNA055/L10 leader	OLEC521	F	CAAGTCCTAGTAGAAAGCAA		55	probe
	OLEC522	R	CACCTAAACCTAGACACGA	269 bp		
SpyRNA056	OLEC523	F	CTGCCTTATAAACCTTTTAA		55	probe
	OLEC524	R	AATGAGTGGCTAGTTATCTC	206 bp		
SpyRNA057	OLEC525	F	AAATCGCCGATGTCTTAGGT		55	probe
	OLEC526	R	TGCAGTAGCAACTGTCAACC	228 bp		
SpyRNA058	OLEC527	F	TCCGGTTTTAAAAGGGTCAG		55	probe
	OLEC528	R	GGAGACAAAAGGCAGGGA	416 bp		
SpyRNA059	OLEC529	F	TGGACCTCTATGTTATTCC		55	probe
	OLEC530	R	TGAGGTTATCCTAAAACCTGA	248 bp		
SpyRNA060	OLEC531	F	GCACTGACCCCAAAGTTGGC		55	probe
	OLEC532	R	GACAAAGACGAGTTCTGATTA	184 bp		
SpyRNA061	OLEC533	F	AGAGGAGTGGTATGAAAACCT		55	probe
	OLEC534	R	GCAAAGATGAGTATAGCAGG	258 bp		
SpyRNA062	OLEC535	F	CTACTAATTCATAAGGTGCC	219 bp	55	probe

	OLEC536	R	CTTGCAATTCCTCTTAAATTA			
SpyRNA063	OLEC537	F	TCATGACCTATTGGGGTTAC		55	probe
	OLEC538	R	TTGCGTACACAGGACTGACA	320 bp		
SpyRNA064/L13 leader	OLEC539	F	TTTCTCTATAACGAATTCGTA		55	probe
	OLEC540	R	CAGAGCAGCTCTGCCATTCTC	209 bp		
SpyRNA065	OLEC543	F	ATTTTCTCCAATAAAAAGAATC		55	probe
	OLEC544	R	TTCCATATAAAAAGGTGGATCA	173 bp		
SpyRNA066	OLEC591	F	CCTAAAAAGCTTGCAATGA		55	probe
	OLEC592	R	AATAGGCTAGCTCCTCAA	147 bp		
SpyRNA067	OLEC593	F	AAGTGTTAAGTCGTCAAAGTA		55	probe
	OLEC594	R	ATAAAAAGACCAAGTGTCAC	188 bp		
SpyRNA068	OLEC595	F	AACAATTAACAATGGCAC		55	probe
	OLEC596	R	CAAAGCAACCACAGATAA	232 bp		
SpyRNA069	OLEC597	F	TCTCCTTCTGAATTCCTAATC		55	probe
	OLEC598	R	AGTAATGATTACTAGGCCG	274 bp		
SpyRNA070	OLEC599	F	TCTCCTTTTTCAAGTTATGA		55	probe
	OLEC600	R	CCTCTTTTTCACATTACCTA	365 bp		
SpyRNA071	OLEC601	F	ATTTCTTTCAATGCAATC		55	probe
	OLEC602	R	TTTAAGAACAAGAAGCGC	130 bp		
SpyRNA072	OLEC603	F	TCCTATTAGATCTTCCACC		55	
	OLEC604	R	CAGAAAGGCAACTCATTA	328 bp		probe
SpyRNA073	OLEC605	F	TTCGTTTTTATCCCTCTTA		55	
	OLEC606	R	GGGTTTATCTAAAAGTCAAA	271 bp		probe
SpyRNA074	OLEC607	F	AACCAACCACTCCCTTC		55	
	OLEC608	R	CGCTCTTTATAAAAAGAGTGC	127 bp		probe
SpyRNA075	OLEC609	F	GAGTGCAAAACTAAAATAACC		55	probe
	OLEC610	R	CCTCTTTTCAGGAGTTATTA	239 bp		
SpyRNA076	OLEC611	F	AACCAGTTATTAACCAGTTA		55	probe
	OLEC612	R	AACCAATTTGGTTCTCTC	176 bp		
purine	OLEC437	F	TGATATGGAGTTAGCTAGGA		55	probe
	OLEC438	R	CAAAGCCAGAACTATAAGA	267 bp		
PYRF	OLEC709	F	ATAATAGTTTAAATTTAGACA		55	probe
	OLEC710	R	CTATAATGTGTCAATATCTCA	153 bp		
SRP	OLEC766	R	AAAGAGCACACCCGAAAAAT		55	Seq./PE**

	OLEC431	F	TGTTTCATCATAAACTCCCAT		55	probe
	OLEC432	R	CCTCTTAGCCTAAATAAAAA	392 bp		
glycine	OLEC768	R	TAATAAGGTCTCTCCAGAAGT		55	PE**
	OLEC435	F	TTTTTGAGGTATGATAGTCC		55	probe
	OLEC436	R	ATTTAACGAGTGCTATCATC	351 bp		
RNaseP	OLEC375	F	TAGCTTAACCTATTATGCAA		55	probe
	OLEC376	R	AGGGATGTGCATACACATTA	290 bp		
	OLEC697	R	CACTTTCAGAAGTACTTTGGC		55	Seq./PE**
	OLEC793	F	CGACTCGTGTGCTCAATCA	281 bp		
tmRNA	OLEC448	F	TTTGGGGTTGTTACGGATTC		55	probe
	OLEC449	R	AACATATTTGTCTACGTCCA	308 bp		
	OLEC693	R	TGAGACTTCTAACAGGACACA		55	Seq./PE**
	OLEC792	F	TAATCATACCATAAGCTGTA	258 bp		
ScpA014	OLEC554	F	TTTCCTGCAGGGTTTTGA		55	probe
	OLEC555	R	GGAGGTCACAAACTAAGCA	637 bp		
ScpA	OLEC557	F	TGGTGCTTCCTCAGAAACCGCTGTTGGTTG		55	Seq./PE**
	OLEC513	R	CATTTAATCATTATGACAAAA	360 bp		
	OLEC558	F	ATTGTCATAGATCAAGACTCC		55	probe
	OLEC559	R	ACCTCAGCTGGTAATGATA	384 bp		
TPP/SpyRNA044	OLEC476	F	CGTCTTTGCTTATTAAC		55	probe
	OLEC477	R	AGGTAGAACGAGTCAATTTA	297 bp		
SPy1212RNA027	OLEC725	F	AAGCAG TTGTCGGAACAATTA ACTT	371 bp	55	probe
SPy1212	OLEC734	F	GCTATCATCAGATTACGCCAAGAGA		55	probe
	OLEC735	R	TCGGATATCTACTCCACGTTCTGAT	520 bp		
SPy2188RNA034	OLEC726	R	AAGGAATACCGATTTTATCAGCAAC		55	probe
SPy2188	OLEC730	F	AAGTTAGAGAAAATCGCTGAGCGA		55	probe
	OLEC731	R	CTCTGCAAAAACA ACTTCTGCCT	473 bp		
SPy1333RNA042	OLEC723	F	GGTATTTAATGTGGCATTGGAGAA		55	probe
	OLEC724	R	CTTTCCAGTTTCAGCATCACGA	508 bp		
SPy1333	OLEC732	F	CATCAGCTAAACCGAAAATTGGT		55	probe
	OLEC733	R	TCGTCTGTTTTAGCGAGTAATTCTG	484 bp		
SPy0516	OLEC655	F	AACAGGCATTCCCTTGAAAA		55	probe
	OLEC656	R	AAATAAGGATTGCCATGAGCA	426 bp		
SPy1211	OLEC662	F	AAAACGCAAGTGAACCACAA	226 bp	55	probe

	OLEC663	R	TGACACGTGGCGACAAAA			
049STagFPT7	OLEC854	F	TAATACGACTCACTATAGGAGTATTAAGTATTGTTTTATGGC		55	STag
049STag1R	OLEC853	R	TTGCGGGCAGAAGTCCAAATGCGATCCATGATAAAACAAGCGTT	283 bp		
6SSTagFPT7	OLEC851	F	TAATACGACTCACTATAGGTACAGAAACATTGCTGTGGCATA		55	STag
6SSTag1R	OLEC852	R	TGCGGGCAGAAGTCCAAATGCGATCCAGAATGCTGTATTGGCA	265 bp		
STAG2	OLEC829	R	GGATCCGACCGTGGTGCCTTTCGCGGCAGAAGTCCAAATGCGATCC			STag
L21	OLEC721	F	TGGTGGTTTAGCATCTAACGA		55	probe
	OLEC722	R	ACTCGCCTGGAATCGTGAGAC	163 bp		
alaS T box	OLEC811	F	GTTATGCTGAAGCTAGTA		55	probe
	OLEC812	R	GCTACTTAACTTTACAAC	232 bp		
glyQS T box	OLEC813	F	CAATCGTTTTTTTGTCA		55	probe
	OLEC814	R	ATAAATTAGGGTGGAAACCGC	140 bp		
pheST T box	OLEC815	F	TTGAATAAGTGTCTATGAGA		55	probe
	OLEC816	R	CGTTATCAGACCTTAATTAA	131 bp		
serST T box	OLEC940	F	ACTTTTTTTCATTTCAGAGA		55	probe
	OLEC941	R	ATGCTGGAGCGCTTTGGC	142 bp		
5S rRNA	OLEC287	F	AGTTAAGTGACGATAGCCTAG		55	probe
	OLEC288	R	CTAAGCGACTACCTTATCTCA	117 bp		
16S rRNA	OliRN242	F	CGGTAACCTAACAGAAAGGG		55	probe
	OliRN243	R	CGTTGTACCAACCATTGTAGC	769 bp		
fasX	OLEC182	F	GAGCAATAACATTTTAGG		55	control
	OLEC183	R	TTACAATCAGCTGATGTG	253 bp		
Cds49KODwFSacI	OLEC821	F	GACAGCGAGCTCATAGAATGATAACAAAATAAA		55	K.O.
Cds49KODwREcoRI	OLEC822	R	AAGTGCGAATTCATCAGCATATTGATCTCCAAT	888 bp	55	K.O.
Cds49KOUppFstI	OLEC823	F	TAGACTCTGCAGGACTGATATCATGGCAGGTAT		55	K.O.
Cds49KOUppRSacI	OLEC824	R	CTGTAAGAGCTCTTCTATTCTACTCTGACTGCA	1098 bp	55	K.O.

\* Annealing Temperature

\*\* Sequencing/Primer Extension

Underlined sequences represent restriction sites. Italic letters represent sequences that are not part of the SF370 genome.

Table 12. Tested sRNA predictions

Candidate	Adjacent genes				Algorithm, predicted 5' and 3' end (length, nt)					
	5' gene		3' gene		sRNA Dir.	Coordinates	QRNA	RNAz	Alifoldz	sRNAPredict2
	gene	Dir.	Gene	Dir.						
SpyRNA001	SPy0041	>	AdhA	>	?	58819-59012		58819-58973 (155)	58859-59012 (154)	
SpyRNA002	<i>pbp1b</i>	>	<i>RpoB</i>	>	?	93328-93484	93406-93455 (50)	93328-93484 (157)	93328-93483 (156)	
SpyRNA003	slo	>	<i>SPy0168</i>	>	?	153316-153409		153316-153409 (94)		
SpyRNA004	<i>SPy0272</i>	>	<i>Fus</i>	>	?	236788-237155	236822-237103 (281)	236899- 237155 (256)	236899-237154 (255)	236788-237098 (311)
SpyRNA005/yybP-ykoY	<i>SPy0622</i>	>	<i>PacL</i>	>	>	500034-500173		500074-500173 (99)	500034-500172 (138)	
SpyRNA006	Ser tRNA	>	SPy0019	>	?	29652-29801		29652-29801 (150)		
SpyRNA007	<i>SPy0019</i>	>	prsA.2	>	?	31958-31995		31958-31995 (38)		
SpyRNA008	<i>SPy1124</i>	<	<i>SPy1125</i>	>	?	921473-921533		921473-921533 (61)		
SpyRNA009	<i>scl</i>	<	<i>SPy1984</i>	>	?	1652875-1653000		1652875-1653000 (126)		
SpyRNA010	purN	>	<i>Spy0031</i>	<	?	43641-45122		43641-45122 (1482)		
SpyRNA011/TPP	<i>SPy0600</i>	<	<i>SPy0601</i>	<	<	482857-483009				482857-483009 (188)
SpyRNA012/PyrR1	<i>SPy0827</i>	>	<i>PyrR</i>	>	>	680165-680331		680165-680331 (167)	680165-680330 (166)	680191-680323 (133)
SpyRNA013					?					
SpyRNA014	<i>scpA</i>	<	<i>SPy2013</i>	>		1680455-1680881	1680767-1680792 (26)			1680455-1680881 (427)
SpyRNA015	<i>SPy2066</i>	>	<i>GroEL</i>	<	?	1722505-1722635				1722505-1722635 (131)
SpyRNA016/L20 leader	<i>cmk</i>	>	<i>InfC</i>	>	>	661385-661541	661399- 661525 (126)	661425-661541 (117)	661425-661540 (116)	661385-661511 (127)
SpyRNA017	<i>sic</i>	<	<i>Emm</i>	<	?	1683605- 1683769	1683639-1683769 (131)			1683605-1683755 (151)
SpyRNA018	<i>csp</i>	<	<i>Cys tRNA</i>	<	?	1728281-1728576				1728281-1728576 (296)
SpyRNA019	<i>hsdM</i>	>	<i>SalR</i>	<	?	1590732-1591158	579678-579757 (79)	1591042-1591158 (117)		1590732-1591081 (350)
SpyRNA020	<i>pyrG</i>	<	<i>RpoE</i>	<	?	1571050-1571194		1571050-1571194 (145)	1571050-1571193 (144)	1571085-1571178 (94)
SpyRNA021/ileS T box	<i>ileS</i>	<	<i>divIVAS</i>	>	?	1245005-1245232		1245067-1245201 (135)	1245067-1245200 (134)	1245005-1245232 (228)
SpyRNA022/L19 leader	<i>SPy0723</i>	>	<i>RplS</i>	>	>	579642-579760		579642-579757 (116)	579642-579760 (119)	579678-579759 (82)
SpyRNA023	<i>mur1.2</i>	>	<i>SPy0861</i>	<	?	709812-709904	709854-709904 (51)			709812-709889 (78)
SpyRNA024	<i>acoL</i>	>	<i>HylA</i>	<	?	840621-840784	840624-840660 (37)			840621-840784 (164)

Candidate	Adjacent genes				Algorithm, predicted 5' and 3' end (length, nt)					
	5' gene		3' gene		sRNA Dir.	Coordinates	QRNA	RNAz	Alifoldz	sRNAPredict2
	gene	Dir.	Gene	Dir.						
SpyRNA025	<i>fhuA</i>	<	<i>MurE</i>	<	?	330502-330595				330502-330595 (94)
SpyRNA026	<i>SPy0867</i>	>	<i>Fms</i>	>	?	717111-717532	717203-717239 (37)	717415-717532 (118)	717415-717531 (117)	717111-717510 (400)
SpyRNA027	<i>Spy1211</i>	<	<i>SPy1212</i>	<	<	996002-996175		996047-996148 (102)	996047-996147 (101)	996002-996175 (174)
SpyRNA028	Ser tRNA	>	<i>SPy0019</i>	>	?	30276-30455		30276-30455 (180)		
SpyRNA029	<i>eutD</i>	>	<i>SPy1131</i>	<	?	924927-925051		924934-925051 (118)		924927-925018 (92)
SpyRNA030	<i>SPy0290</i>	>	<i>DacA</i>	<	?	254688-254834		254688-254834 (147)		
SpyRNA031	tmRNA	>	<i>SPy1290</i>	>	?	1065515-1065622				1065515-1065622 (108)
SpyRNA032	16S rRNA	<	<i>SPy1613</i>	<	?	1336816-1337022				1336816-1337022 (207)
SpyRNA033	<i>SPy0208</i>	>	<i>SPy0210</i>	<	?	187780-188123				187780-188123 (344)
SpyRNA034	<i>SPy2188</i>	<	<i>SdhB</i>	>	<	1822330-1822555				1822330-1822555 (226)
SpyRNA035	<i>SPy2186</i>	>	<i>SPy2188</i>	>	?	1822330-1822555				1822330-1822555 (226)
SpyRNA036	<i>greA</i>	>	<i>SPy0351</i>	>	?	306646-307025				306646-307025 (379)
SpyRNA037	<i>ftsA</i>	<	<i>DivIB</i>	<	?	1251191-1251270				1251191-1251270 (80)
SpyRNA038	<i>SPy1942</i>	<	<i>CysE</i>	<	?	1616407-1616689				1616407-1616689 (283)
SpyRNA039	<i>SPy0208</i>	>	<i>SPy0210</i>	<	?	188359-188432				188359-188432 (74)
SpyRNA039	<i>prgA</i>	>	<i>RpsL</i>	>	?	235748-235883		235748- 235867 (120)	235767- 235883 (117)	
SpyRNA040/trsA T box	<i>trsA</i>	<	<i>Spy2209</i>	>	<	1842020-1842331				1842020-1842331 (312)
SpyRNA041	<i>SPy1608</i>	>	<i>SPy1610</i>	>	?	1329527-1329649				1329527-1329649 (123)
SpyRNA042	<i>SPy1333</i>	<	<i>SPy1335</i>	>	<	1106399-1106788	1106731-1106778 (48)	1106730- 1106788 (59)	1106730- 1106787 (58)	1106399-1106631 (233)
SpyRNA043	<i>SPy0290</i>	>	<i>DacA</i>	<	?	253870-253979				253870-253979 (110)
SpyRNA044	<i>acoL</i>	>	<i>HylA</i>	<	?	840292-840392				840292-840392 (101)
SpyRNA045	Gln tRNA	>	<i>RpsA</i>	>	?	758560-758690	758597-758690 (94)			758560-758658 (99)
SpyRNA046/valS T box	<i>SPy1570</i>	<	<i>SPy1571</i>	<	<	1295632-1295860			1295716-1295843 (128)	1295632-1295860 (229)
SpyRNA047	<i>SPy2191</i>	<	<i>SPy2193</i>	<	?	1825347-1825568		1825347-1825484 (138)		1825491-1825568 (78)

Candidate	Adjacent genes				Algorithm, predicted 5' and 3' end (length, nt)					
	5' gene		3' gene		sRNA Dir.	Coordinates	QRNA	RNAz	Alifoldz	sRNAPredict2
	gene	Dir.	Gene	Dir.						
SpyRNA048	<i>rplL</i>	>	<i>SPy1075</i>	>	?	883502-883729	883561-883632 (72)			883502-883729 (228)
SpyRNA049	<i>SPy1044</i>	>	<i>SPy1046</i>	>	<	854268-854668	854268-854527 (159)			854317-854668 (352)
SpyRNA050/thrS T box	<i>SPy0516</i>	>	<i>ThrS</i>	>	>	416933-417175	417044- 417103 (59)			416933-417175 (243)
SpyRNA051	<i>glyS</i>	<	<i>GlyQ</i>	<	?	1403469-1403535				1403469-1403535 (67)
SpyRNA052	<i>SPy0159</i>	<	<i>PurA</i>	>	?	145555-145712			145595-145709 (115)	145555-145712 (158)
SpyRNA053	<i>SPy1927</i>	>	<i>RpsI</i>	<	?	1608735-1609187				1608735-1609187 (453)
SpyRNA054	<i>tnpA</i>	<	<i>SPy0567</i>	>	?	456755-456907				456755-456907 (153)
SpyRNA055/L10 leader	<i>trmE</i>	<	<i>RplJ</i>	>	>	882351-882520	882410- 882511 (101)	882399-882516 (118)		882351-882520 (170)
SpyRNA056	<i>deaD</i>	<	<i>PrfC</i>	<	?	1175520-1175698			1175520-1175634 (115)	1175540-1175698 (159)
SpyRNA057	<i>SPy0949</i>	<	<i>SPy0952</i>	>	?	785852-786006				785852-786006 (155)
SpyRNA058	<i>Arg tRNA</i>	>	<i>SPy0726</i>	>	?	580439-580871	580439-580463 (25)			580442- 580871 (430)
SpyRNA059	<i>tnpA</i>	>	<i>SPy0567</i>	<	?	519005-519094				519005-519094 (90)
SpyRNA060	<i>oppF</i>	>	<i>SPy0299</i>	<	?	263371-263445	263409-263445 (37)			263371-263445 (75)
SpyRNA061	<i>SPy0486</i>	>	<i>SPy0488</i>	>	?	394232-394559	394371- 394403 (32)			394232-394559 (328)
SpyRNA062	<i>SPy1725</i>	<	<i>Ser tRNA</i>	<	?	1430148-1430242				1430148-1430242 (95)
SpyRNA063	<i>SPy2166</i>	>	<i>SPy2169</i>	>	?	1803129-1803393				1803129-1803393 (265)
SpyRNA064/L13 leader	<i>rplM</i>	<	<i>SPy1934</i>	<	<	1610039-1610177	1610039-1610158 (120)	1610048-1610178 (131)	1610048-1610177 (130)	
SpyRNA065	<i>groES</i>	<	<i>ClpC</i>	<	?	1724595-1724704		1724595-1724704 (110)	1724595-1724703 (109)	
SpyRNA066	<i>SPy0045</i>	>	<i>RpsJ</i>	>	?	63546-63713		63546-63713 (168)		
SpyRNA067	<i>SPy0181</i>	>	<i>SPy0182</i>	>	?	166761-166864			166761-166864 (104)	
SpyRNA068	<i>SPy1002</i>	>	<i>Lys</i>	>	?	817305-817387		817305- 817387 (83)		
SpyRNA069	<i>fbp</i>	<	<i>SPy1016</i>	>	>	826019-826107	826019-826107 (89)			
SpyRNA070	<i>SPy1180</i>	<	<i>SPy1181</i>	>	?	971324-971377		971324-971377 (54)		
SpyRNA071	<i>SPy1227</i>	<	<i>SPy1228</i>	<	?	1013489-1013632		1013489-1013632 (144)	1013489-1013631 (143)	



Candidate	Adjacent genes				sRNA Dir.	Coordinates	Algorithm, predicted 5' and 3' end (length, nt)			
	5' gene		3' gene				QRNA	RNAz	Alifoldz	sRNAPredict2
	gene	Dir.	Gene	Dir.						
SpyRNA072	<i>SPy1395</i>	<	<i>SPy1398</i>	<	?	1159391-1159482		1159391-1159482 (92)		
SpyRNA073	<i>SPy1508</i>	<	<i>ClpE</i>	>	?	1238451-1238560		1238451-1238560 (110)		
SpyRNA074	<i>hrcA</i>	<	<i>SPy1764</i>	<	?	1460950-1461082		1460970-1461082 (113)	1460950-1461081 (132)	
SpyRNA075	<i>flaR</i>	>	<i>SmeZ</i>	>	?	1664988-1665009	1664988-1665009 (22)			
SpyRNA076	<i>lmb</i>	<	<i>SPy2009</i>	<	?	1675616-1675642	1675616-1675642 (27)			
SRP	<i>SPy0210</i>	<	<i>SpeG</i>	>	>	190315-190477	190318-190433 (115)	190315-190435 (120)	190315-190434 (119)	190318-190477 (160)
FMN	<i>SPy0371</i>	>	<i>SPy0373</i>	>	>	319155-319381		319165-319231 (66)		319155-319381 (226)
L21 leader	<i>SPy0818</i>	>	<i>RplU</i>	>	>	676258-676400	676268-676400 (132)	676258-676392 (135)		
PyrR2	<i>SPy0899</i>	>	<i>PyrF</i>	>	?	744492-744605	744506-744573 (67)	744492-744605 (114)	744492-744604 (113)	
purine	<i>SPy1135</i>	>	<i>Xpt</i>	>	>	930675-930850	930766-930834 (69)	930750-930838 (89)		930675-930850 (176)
glycine	<i>pcrA</i>	<	<i>SPy1270</i>	>	>	1046389-1046615	10466113-10466164 (51)			1046389-1046615 (227)
tmRNA					>	1065083-1065361	1065242-1065361 (120)		1065083-1065319 (237)	
RNase P	<i>SPy1644</i>	<	<i>SPy1646</i>	<	<	1365142-1365566	1365163-1365281 (119)	1365142-1365566 (425)	1365398-1365515 (118)	1365146-1365550 (405)
6S RNA	<i>SPy1992</i>	>	<i>Lys tRNA</i>	>	>	1663528-1663821			1663712-1663821 (110)	
<i>pheST</i> T box <sup>a</sup>	<i>endA</i>	>	<i>PheS</i>	>	>	623720-623873			623720-623873 (154)	
<i>alaS</i> T box <sup>a</sup>	<i>alaS</i>	<	<i>PrsA</i>	<	<	1152857-1152975		1152857-1152975 (119)	1152857-11529754 (118)	
<i>glyQS</i> T box <sup>a</sup>	<i>glyQ</i>	<	<i>SPy1691</i>	<	<	1403469-1403535		1404666-1404820 (155)		
<i>serST</i> box <sup>a</sup>	<i>SPy1741</i>	>	<i>SerS</i>	>	>	1443423-1443461	1443423-1443461 (38)			

All positive candidates have a coloured background: positive tested predictions of the Rfam database (violet), T boxes (turquoise) and novel sRNA candidates (light blue).

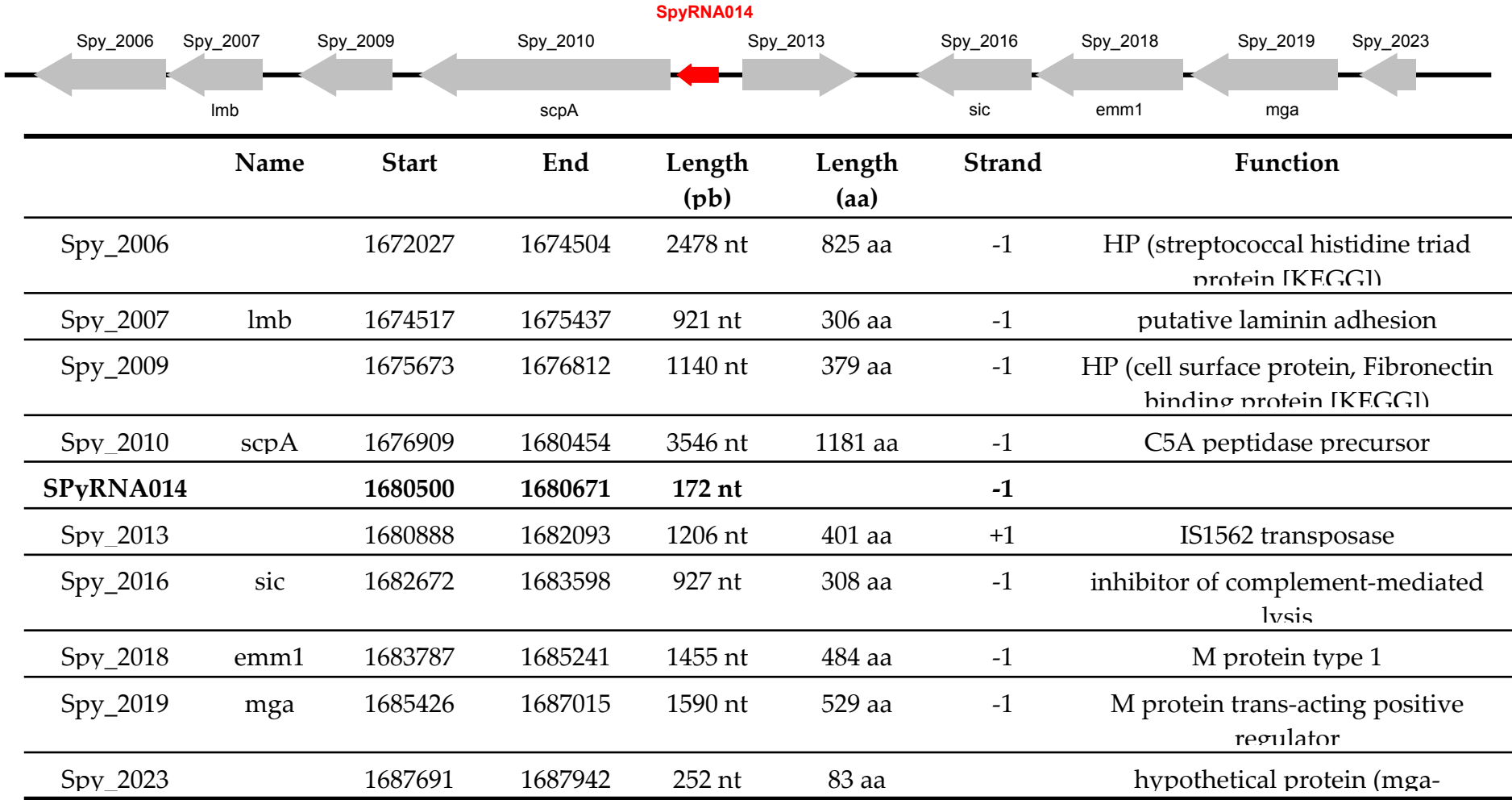


Fig. 21. Annotation of genes surrounding SpyRNA014.

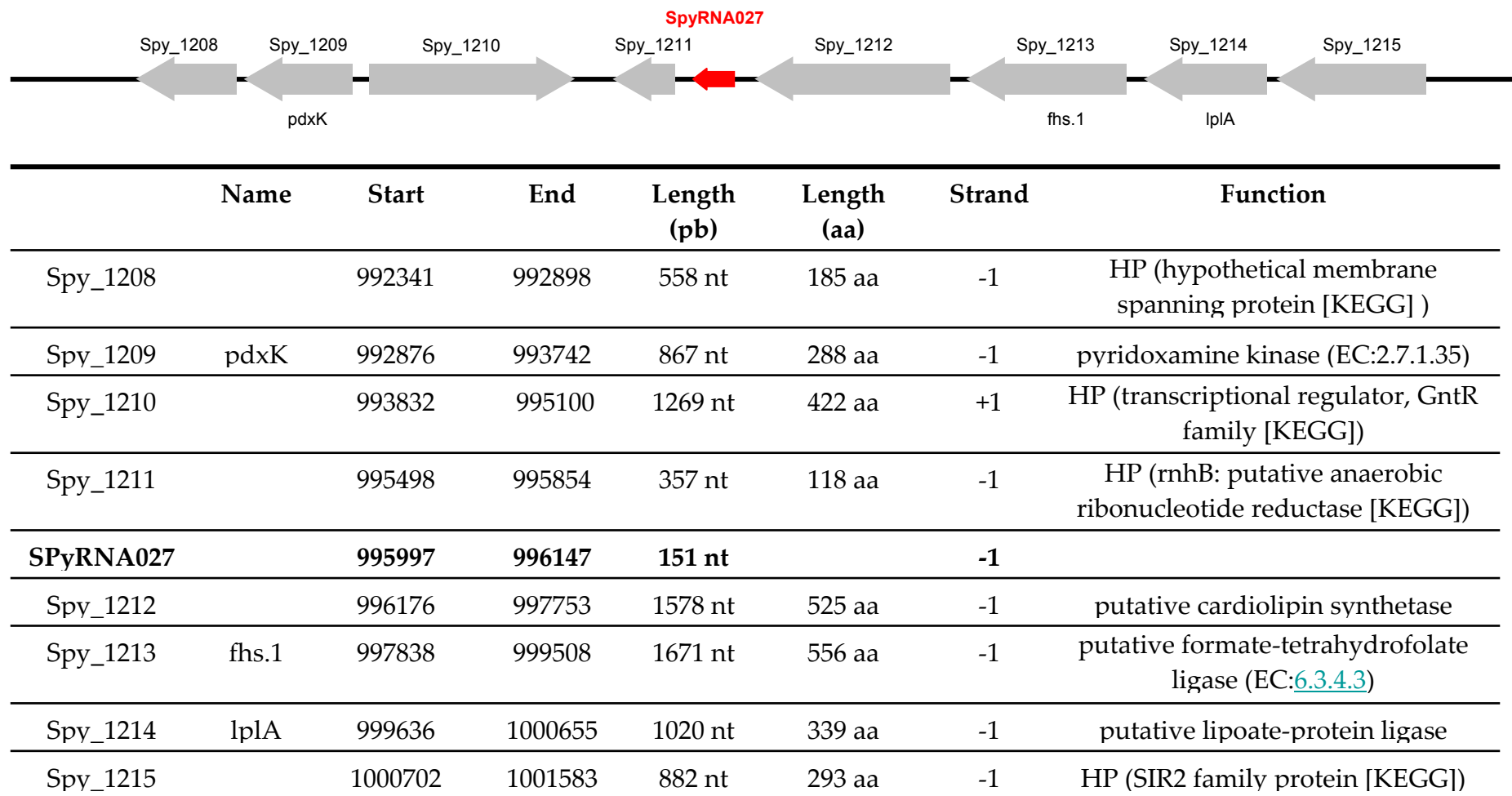
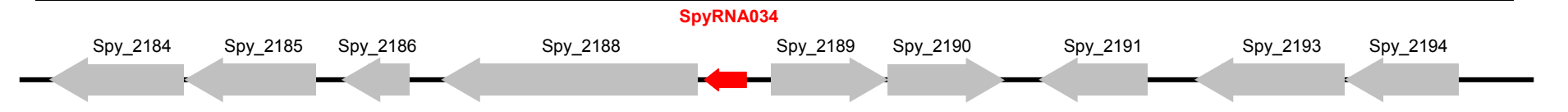
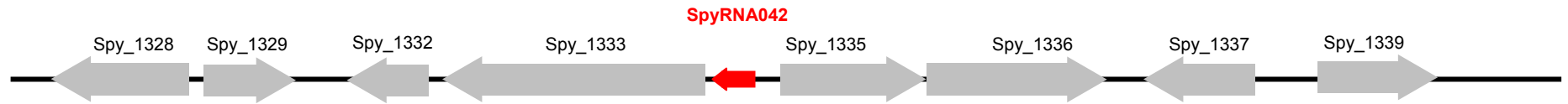


Fig. 22. Annotation of genes surrounding SpyRNA027.



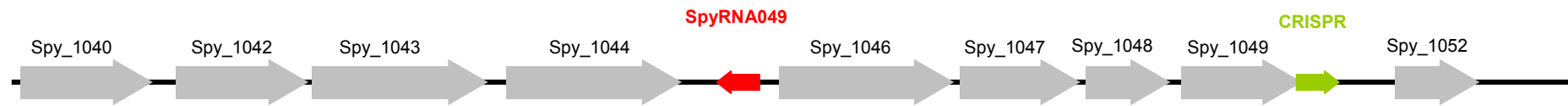
Name	Start	End	Length (pb)	Length (aa)	Strand	Function
Spy_2184	1816020	1817996	1977 nt	658 aa	-1	HP (phosphoesterase, DHH family protein [KEGG & NCBI Blast])
Spy_2185	1818087	1819985	1899 nt	632 aa	-1	tRNA uridine 5-carboxymethylaminomethyl modification enzyme
Spy_2186	1820109	1820426	318 nt	105 aa	-1	HP (phosphohydrolase [KEGG & NCBI Blast])
Spy_2188	1821203	1822324	1122 nt	373 aa	-1	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase
<b>SPyRNA034</b>	<b>1822345</b>	<b>1822473</b>	<b>129 nt</b>		<b>-1</b>	
Spy_2189	1822622	1823293	672 nt	223 aa	+1	putative L-serine dehydratase beta subunit
Spy_2190	1823305	1824177	873 nt	290 aa	+1	putative L-serine dehydratase alpha subunit (EC:4.3.1.17)
Spy_2191	1824590	1825204	615 nt	204 aa	-1	HP (transglycosylase SLT domain family protein [KEGG])
Spy_2193	1825576	1826376	801 nt	266 aa	-1	HP (cbiQ: cobalt transport protein [KEGG])
Spy_2194	1826369	1827211	843 nt	280 aa	-1	Putative cobalt/nickel ABC transporter (ATP-binding protein)

**Fig. 23.** Annotation of genes surrounding SpyRNA034.



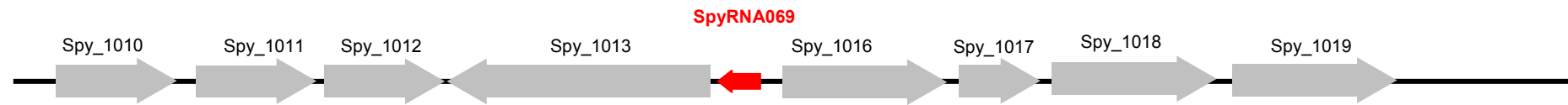
	Name	Start	End	Length (pb)	Length (aa)	Strand	Function
Spy_1328	bglA.2	1102494	1103894	1401 nt	466 aa	-1	putative beta-glucosidase (EC: <a href="#">3.2.1.21</a> )
Spy_1329		1104119	1104616	498 nt	165 aa	+1	HP (nicotinamide mononucleotide transporter [KEGG & NCBIblast])
Spy_1332		1104820	1104978	159 nt	52 aa	-1	HP
Spy_1333	obgE, cgtA, obg, yhbZ	1105047	1106354	1308 nt	435 aa	-1	GTPase ObgE
<b>SPyRNA042</b>		<b>1106393</b>	<b>1106572</b>	<b>180 nt</b>		<b>-1</b>	
Spy_1335		1106787	1107302	516 nt	171 aa	+1	conserved hypothetical protein - transposon IS861 associated
Spy_1336		1107323	1108129	807 nt	268 aa	+1	putative transposase - IS861 associated
Spy_1337	rsuA	1108178	1108912	735 nt	244 aa	-1	putative 16S pseudouridylate synthase
Spy_1339		1109024	1109389	366 nt	121 aa	+1	HP

Fig. 24. Annotation of genes surrounding SpyRNA042.



	Name	Start	End	Length (pb)	Length (aa)	Strand	Function
	Spy_1040	850988	852118	1131 nt	376 aa	+1	coproporphyrinogen III oxidase (EC: <a href="#">1.3.3.3</a> )
	Spy_1042	852312	852881	570 nt	189 aa	+1	HP (acyl-acyl carrier protein thioesterase [KEGG & NCBIblast])
	Spy_1043	852881	853645	765 nt	254 aa	+1	HP (4-nitrophenylphosphatase (EC: <a href="#">3.1.3.41</a> ) [KEGG & NCBIblast])
	Spy_1044	853645	854277	633 nt	210 aa	+1	HP (hypothetical membrane spanning protein [KEGG & NCBIblast])
	<b>SPyRNA049</b>	<b>854376</b>	<b>854546</b>	<b>171 nt</b>		<b>-1</b>	
	Spy_1046	854757	858863	4107 nt	1368 aa	+1	CRISPR-associated protein Csn1
	Spy_1047	858863	859732	870 nt	289 aa	+1	CRISPR-associated protein Cas1
	Spy_1048	859729	860070	342 nt	113 aa	+1	CRISPR-associated protein Cas2
	Spy_1049	860060	860722	663 nt	220 aa	+1	CRISPR-associated protein Csn2
	Spy_1052	861369	861776	408 nt		+1	pseudogene

Fig. 25. Annotation of genes surrounding SpyRNA049.



	Name	Start	End	Length (pb)	Length (aa)	Strand	Function
Spy_1010	mutX	821203	821677	477	158 aa	+1	putative 7,8-dihydro-8-oxoguanine-triphosphatase
Spy_1011		821737	822916	1179	393 aa	+1	HP (hypothetical membrane spanning protein [KEGG])
Spy_1012		822908	824153	1245	415 aa	+1	HP (tetratricopeptide repeat family protein [KEGG])
Spy_1013	fbp54	825867	824214	1653	550 aa	-1	putative fibronectin-binding protein-like protein A
<b>SPyRNA069</b>		<b>825970</b>	<b>826157</b>	188		<b>-1</b>	
Spy_1016		826220	827216	996	332 aa	+1	HP (ABC transporter substrate-binding protein [KEGG])
Spy_1017		827223	827385	162	54 aa	+1	hypothetical protein
Spy_1018		827564	828431	867	289 aa	+1	putative ABC transport protein (permease)
Spy_1019		828430	829186	756	252 aa	+1	putative ABC transporter (ATP-binding protein)

**Fig. 26.** Annotation of genes surrounding SpyRNA069.

## 9. Appendix B

---

Manuscript submitted on July 22, 2008 to the journal "Antimicrobial Agents and Chemotherapy".

### **Novel pleuromutilin derivatives and their interactions with *Escherichia coli* and *Staphylococcus aureus* ribosomes**

Susanne Paukner<sup>1</sup>, Christine Schneckenleithner<sup>2</sup>, Markus Mair<sup>1,2</sup>, Doris Veit<sup>2</sup>, Agnieszka Kokot<sup>2</sup>, Daniel N. Wilson<sup>3</sup>, Rodger Novak<sup>1</sup> and Emmanuelle Charpentier<sup>2\*</sup>

Nabriva Therapeutics AG, Leberstrasse 20, A-1112 Vienna, Austria<sup>1</sup>; Max F. Perutz Laboratories, Department of Microbiology and Immunobiology, University Vienna, Dr. Bohrgasse 9, A-1030 Vienna, Austria<sup>2</sup>; Gene Center and Department for Chemistry and Biochemistry, University of Munich, Feodor-Lynenstr. 25, D-81377 Munich, Germany<sup>3</sup>

**Running title:** Activities and ribosomal interaction sites of novel pleuromutilin derivatives in the human pathogen *Staphylococcus aureus*

**Key words:** *Staphylococcus aureus*, pleuromutilin, footprinting, ribosome, peptidyl transferase center, 23S rRNA, antibiotic ribosome interaction

#### **Correspondent footnote:**

\*Corresponding author. Mailing address: Max F. Perutz Laboratories, University of Vienna, Dr. Bohrgasse 9/4, 1030 Vienna, Austria. Phone: (43) (1) 4277 54601. Fax: (43) (1) 4277 9546. E-mail: [emmanuelle.charpentier@univie.ac.at](mailto:emmanuelle.charpentier@univie.ac.at)



**ABSTRACT**

Pleuromutilin antibiotics are protein synthesis inhibitors that are currently in clinical development for the treatment of respiratory tract and skin infections in humans. In this study, we report two novel chemically designed pleuromutilin derivatives, BC-3291 and BC-3311. Compared to other pleuromutilins and BC-3291, BC-3311 exhibited superior *in vitro* antibacterial activities against a number of clinical isolates including staphylococcal and streptococcal species commonly associated with respiratory and skin infections. To understand the molecular basis that could explain the improved activities of BC-3311, the interaction sites of different derivatives with the target 23S rRNA of both *Escherichia coli* and *Staphylococcus aureus* ribosomes were determined using chemical footprinting. Analysis of the footprints with *S. aureus* ribosomes revealed protections at nucleotides A2058, A2059, U2584 and U2585, which were also detected with *E. coli* ribosomes. Remarkably, in contrast to footprints in *E. coli*, a lack of reactivity at nucleotide U2506, an enhancement at nucleotide C2192 and protections at nucleotides U2533 and U2563 were observed in *S. aureus*, thus suggesting that the conformation of *S. aureus* ribosomes slightly differs from that of *E. coli* ribosomes. Molecular modeling of the footprinting data supports the proposal that distinct conformational changes of the pleuromutilin molecules adopted upon binding to the target and bacterial species-specific conformations of the pleuromutilin binding pocket contribute to explain differences in antibacterial activities of the pleuromutilin derivatives. Taken together, our work emphasizes the need to study the interactions of pleuromutilins with the ribosomes of target bacterial pathogens using a multidisciplinary approach.

## INTRODUCTION

Respiratory tract infections (RTIs) and skin and skin structure infections (SSSIs) are among the most prevalent bacterial infections worldwide . These infections can cause considerable morbidity and mortality with a major impact on public health . In recent years, the management of these infections has been challenged by the escalation of antimicrobial resistance in many pathogens . Among others, the treatment of infections caused by methicillin-resistant *Staphylococcus aureus* (MRSA) represents a challenge to the health care system and physicians . In a search for novel antibacterial agents, pleuromutilin antibiotics have benefited from a renewed interest due to their potent activity against resistant pathogens including MRSA.

Pleuromutilins define a class of protein synthesis inhibitors derived from the natural product pleuromutilin of the basidiomycete *Pleurotus mutilus* (now *Clitopilus scyphoides*) . Of the semi-synthetic derivatives, tiamulin and valnemulin (Econor<sup>®</sup>) have been used almost exclusively in veterinary medicine to treat swine dysentery caused by *Brachyspira spp.* and enzootic pneumonia caused by *Mycoplasma spp.* Recent efforts in medicinal chemistry have resulted in the development of novel pleuromutilin derivatives suitable for human therapy . Among them, retapamulin has recently been approved for human use as a topical agent (Altabax<sup>®</sup> or Altargo<sup>®</sup>) for short-term treatment of superficial skin infections caused by *Streptococcus pyogenes* and methicillin sensitive *S. aureus* (MSSA) such as impetigo and infected small lacerations, abrasions or sutured wounds .

The general structure of pleuromutilins consists of the tricyclic mutilin core, which is substituted at the C14 position on the octane ring with various side-chains. The diverse chemical nature of the side-chain results in carbamate derivatives , or sulfanyl-acetate derivatives such as valnemulin, tiamulin or retapamulin (Fig. 1A, 1B). Previous studies suggested that variations in chemical derivations of the side-chain constitute the basis for differences in activity among pleuromutilin derivatives or their lack of cross-resistance with other antibiotics . Pleuromutilin antibiotics are potent inhibitors of protein synthesis. Early studies showed that the pleuromutilin derivative tiamulin hinders peptide bond formation by targeting specifically tRNA substrate binding at both the acceptor (A-) and donor (P-) sites of the peptidyl transferase centre (PTC) of the *E. coli* 50S large ribosomal subunit . Recent biochemical and crystallographic studies support the

earlier data and show in addition that the tricyclic mutilin core of pleuromutilins is positioned similarly in a tight pocket at the A-tRNA binding site while the C14-extensions point toward the P-tRNA binding site . Interaction sites were analyzed and identified as nucleotides, which are located in the central loop region of domain V of 23S rRNA and have been implicated in part in the dynamically mobile PTC . Altogether the current model suggests that interactions of pleuromutilins with the rRNA in the PTC pocket would prevent the correct positioning of the CCA-ends of tRNAs for peptide transfer . However, because knowledge of the dynamics of the protein synthesis process is still lacking, the exact mechanism of inhibition by pleuromutilins remains poorly understood.

Although pleuromutilin derivatives show activity against gram-positive pathogens, the interaction sites of the antibiotics with the ribosomal target have been almost exclusively determined with ribosomes from *E. coli* (chemical footprinting experiments) or *Deinococcus radiodurans* (crystal structures) . Only one study reports the analysis of interaction binding sites of tiamulin with its actual target, the ribosome from *Brachyspira hyodysenteriae* . Using chemical probing, the authors show that the interaction sites differ when comparing the *E. coli* and *B. hyodysenteriae* ribosomes. Therefore, although the PTC is considered to be very conserved among bacterial species, molecular differences exist that may promote pleuromutilin discrimination by involvement of an allosteric type of mechanism . Determining the binding sites of pleuromutilins on the ribosome of *S. aureus*, the actual target organism of pleuromutilin derivatives for human use, should provide important knowledge for improving rational drug design, e.g. the development of more efficacious agents.

In this study, we focus on two novel pleuromutilin derivatives, BC-3311 and BC-3291. We present the structure of the two derivatives and their *in vitro* antibacterial activities against selected pathogens. Using chemical footprinting, we analyzed their interaction sites at the PTC of both *E. coli* and *S. aureus* ribosomal targets and compared them with the interaction sites of other pleuromutilins, leading to a model for the interaction of the novel derivatives with the *E. coli* and *S. aureus* PTCs. This is the first study describing biochemically revealed binding sites of pleuromutilins on the ribosomal target from a gram-positive bacterial pathogen.

## MATERIALS AND METHODS

**Bacterial strains and culture conditions.** Bacterial strains were purchased from the American Type Culture Collection (ATCC) and the National Collection of Type Cultures (NCTC) or were clinical isolates obtained from the General Hospital in Vienna (AKH, Vienna, Austria), I. Chopra (University of Leeds, Leeds, UK), F.J. Schmitz (SENTRY study, Klinikum Minden, Minden, Germany), Novartis (NIBR, Vienna, Austria) or the Clinical Microbiology Department (University College London Hospitals, London, UK) (Table 1). The identity of all strains was confirmed using the BBL Crystal™ Identification Systems (Beckton Dickinson). For the determination of antibiotic MIC (minimum inhibitory concentration), all strains were maintained as recommended by the Clinical and Laboratory Standards Institute (CLSI, former NCCLS) guidelines. For the preparation of ribosomes, *E. coli* MRE600 and *S. aureus* NCTC8325 were cultured as indicated below.

**Antimicrobial agents.** All antibiotics were provided by Nabriva Therapeutics AG (Vienna, Austria). Stock solutions of all compounds were prepared according to the CLSI guidelines. Tiamulin and valnemulin were dissolved in deionized water, and azamulin, BC-3291, BC-3311, linezolid and pleuromutilin were dissolved in dimethyl sulfoxide (DMSO, 10%). For the determination of MICs (minimal inhibitory concentrations), the solutions were further diluted in deionized water and subsequently with culture medium to reach the required drug concentrations (maximal final DMSO concentration: 1%). All antibiotic weights were corrected for salt forms and referred to the pure drug substance.

**Antibacterial susceptibility.** MICs were determined as recommended by the approved standard reference recommendations of the CLSI using the agar dilution technique for staphylococci, enterococci, *Moraxella catarrhalis* and *E. coli* and the microbroth dilution technique for streptococci and *Haemophilus influenzae*. The inoculum was prepared according to the growth method (CLSI) with a final inoculum size of  $\sim 5 \times 10^7$  CFU/ml. MIC determination was performed in duplicate and repeated at least twice.

**Preparation of ribosomes.** *E. coli* ribosomes were prepared as described earlier with some modifications. Briefly, *E. coli* MRE600 was grown in LB (Luria Bertani) broth to the logarithmic growth phase and harvested by centrifugation.

Cell pellets were washed twice with cooled buffer A (20 mM Tris-HCl (pH 7.5), 100 mM NH<sub>4</sub>Cl, 10 mM MgCl<sub>2</sub>, 0.5 mM EDTA and 6 mM 2-Mercaptoethanol) and immediately frozen and stored at -80°C until further use. After thawing on ice, RNase-free DNase I (Roche Diagnostics GmbH) was added and the cells were opened by sonication. After cooling on ice, the lysate was recovered by centrifugation and layered on sucrose buffer D (20 mM Tris-HCl (pH 7.5), 500 mM NH<sub>4</sub>Cl, 10 mM MgCl<sub>2</sub>, 0.5 mM EDTA and 1.1 M sucrose). This was followed by centrifugation for 16 h. The clear 70S ribosomal pellet was rinsed with buffer B (20 mM Tris-HCl (pH 7.5), 500 mM NH<sub>4</sub>Cl, 10 mM MgCl<sub>2</sub> and 0.5 mM EDTA), resuspended in buffer C (50 mM Tris-HCl (pH 7.5), 150 mM NH<sub>4</sub>Cl, 5 mM MgCl<sub>2</sub> and 6 mM 2-mercaptoethanol) and the optical density was then measured at 260 nm (14.4 A<sub>260</sub> units = 1 mg/ml 70S ribosomes or 1 A<sub>260</sub> unit = 24 pmol 70S ribosomes). 60-90 mg of 70S ribosomes were layered on a 10-40% sucrose gradient and centrifuged for 16 h. Fractions containing the 70S ribosomes were collected, the ribosomes were reassembled by addition of 10 mM MgCl<sub>2</sub> and collected by centrifugation for at least 12 h. Ribosomes were further resuspended in buffer A, precipitated with ethanol and ammonium acetate and finally stored at -80°C.

*S. aureus* ribosomes were prepared as follows. *S. aureus* NCTC 8325 was grown in THB (Todd Hewitt Broth, Becton Dickinson) to the logarithmic growth phase and harvested by centrifugation. Cell pellets were washed with buffer A and immediately frozen and stored at -80°C until further use. After thawing on ice, the pellets were resuspended in buffer A containing RNase free DNase I and lysostaphin (Sigma-Aldrich) (to a final concentration of 250 µg/ml) followed by incubation at 37°C for 30 min and sonication. Ribosomes were then prepared following the same procedure as described above for the *E. coli* ribosomes.

For chemical modification, precipitated ribosomes were centrifuged, resuspended in RNase free water and incubated with the antibiotics as described below. For sequencing reactions, precipitated ribosomes were centrifuged, resuspended in sodium acetate/EDTA solution followed by three phenol-chloroform-isoamyl alcohol extractions and two chloroform extractions. The rRNA was then precipitated with ice-cold ethanol and after centrifugation, the rRNA was resuspended in RNase free water.

**Chemical footprinting.** Ribosomes (20 µM) were incubated with 2 to 500 µM of antibiotics in modification buffer (50 mM HEPES-KOH (pH 8), 10 mM MgCl<sub>2</sub>,

100 mM KCl, 5 mM dithiothreitol) for 30 min at 37°C. Modifications were performed essentially as described earlier. DMS (dimethylsulfate) was used to probe the accessibility of the N1 position of adenine and N3 of cytosine and CMCT [1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-p-toluene sulfonate] was used to probe the accessibility of the N3 of uracil and exceptionally N1 of guanine. 200 µl aliquots of the antibiotic-ribosome complexes were modified with either 4 µl of DMS (1:6 dilution in ethanol) for 10 min at 37°C or with 80 µl of CMCT (42 mg/ml in RNase free water) for 20 min at 37°C. DMS reactions were stopped by adding 25 µl stop buffer (1 M Tris-HCl (pH 7.75), 1 mM EDTA and 1 M 2-mercaptoethanol) followed by precipitation. CMCT reactions were stopped by precipitation of the ribosomes with glycogen and ice-cold ethanol. After centrifugation, the ribosomes were resuspended in a sodium acetate/EDTA containing solution and extracted with phenol and chloroform. The rRNA was precipitated with ethanol, resuspended in RNase free water and the modifications were monitored by primer extension analysis on 23S rRNA using AMV reverse transcriptase (Promega, Madison, WI, USA) and 5'-[<sup>32</sup>P]-labeled deoxyoligonucleotide primers (Table 1). The products of cDNA primer extension and sequencing reaction for nucleotide identification were separated on 8% polyacrylamide sequencing gels as described previously. Positions of the stops in cDNA synthesis and intensities of the modifications were visualized and quantified using a PhosphorImager (Variable Mode Imager Typhoon 8600, Amersham Biosciences Europe GmbH, Freiburg, Germany). Chemical footprinting experiments were performed at least three times to confirm reproducibility of the data and allow estimation of significance. Intensities of the radioactive bands were analyzed using PhosphorImager and ImageQuant v5.1 software (Molecular Dynamics).

**Molecular Modeling.** 3D structures for BC-3291, BC-3311 and azamulin were generated using ACD/ChemSketch (Advanced Chemistry Development, Inc., Toronto, ON, Canada, [www.acdlabs.com](http://www.acdlabs.com)). The drugs were modeled onto the ribosome by superposition of the conserved pleuromutilin core with ribosome-bound pleuromutilins. The C14-side-chains were initially placed based on similarity with structurally determined positions of chemically-related pleuromutilins. Specifically, BC-3311 and azamulin were modeled on the basis of SB-280080 (PDB2OGN; ) since this involves simple substitution of the terminal cyclic ring of the C14-side-chain, whereas the extended side-chains of BC-3291

and valnemulin were positioned similarly to the extended side-chain of tiamulin (PDB1XBP; ). All modeling was done using Coot , PyMol (<http://www.pymol.org>) and regularization with CNS .

**Nucleotide numbering.** Nucleotides are numbered according to the sequence of 23S rRNA from *E. coli* to allow a direct comparison with previously described data.

## RESULTS

**Chemical structure of novel pleuromutilin compounds.** A number of pleuromutilin molecules were derived from Nabriva's research and development program. Differences in the structures of the new derivatives reside in the chemical nature of the C14-side-chain-extension. Here, we report the structure of two of these derivatives, BC-3311 and BC-3291. BC-3311 with an azepan-4-ylsulfanyl-side-chain (Fig. 1A) exhibited excellent antibacterial activity and *in vivo* efficacy. BC-3291 with an extended methyl-piperazine-side-chain (Fig. 1B) differs from BC-3311 in the chain length (Fig. 1A, 1B) and antibacterial activity. On this basis, BC-3291 was chosen as comparator in this study together with conventional pleuromutilins (pleuromutilin, tiamulin, azamulin and valnemulin).

***In vitro* antibacterial activities of BC-3311 and comparators.** The antibacterial activities of BC-3311 and BC-3291 were determined against representative strains of the most important bacterial pathogens involved in RTIs and SSSIs including methicillin-sensitive (MSSA) and methicillin resistant *S. aureus* (MRSA), coagulase-negative staphylococci, *S. pyogenes*, *Streptococcus pneumoniae*, *H. influenzae* and *M. catarrhalis* and compared with other antibacterials (Table 2). BC-3311 exhibited potent activity against all analyzed gram-positive pathogens with MICs of  $\leq 0.1$   $\mu\text{g/ml}$  against staphylococci and *S. pyogenes* and MICs of  $\leq 0.2$   $\mu\text{g/ml}$  against *S. pneumoniae* or *Enterococcus faecium*. When compared with other pleuromutilin derivatives and standard antibiotics for human use, BC-3311 was one of the most active compounds against staphylococcal species including MRSA, with MICs (0.05-0.2  $\mu\text{g/ml}$ , MIC<sub>90</sub> of 0.1  $\mu\text{g/ml}$ ) being 4- to 16-fold lower than that of azithromycin or doxycycline against susceptible strains (Table 2) and approximately 20-fold lower than linezolid (Table 3). Remarkably, BC-3311 was fully active against MRSA strains resistant to macrolides, ketolides, lincosamides, tetracyclines, fluoroquinolones and vancomycin (Tables 2, 3). Activities included other gram-positive pathogens like *S. pyogenes*, *E. faecium* and *S. pneumoniae* with activities being comparable or superior to the most active standard antibiotics like azithromycin, moxifloxacin or doxycycline. Although pleuromutilins generally exhibit low activity against gram-negative species, BC-3311 showed potent activity against *M. catarrhalis* (MICs 0.025  $\mu\text{g/ml}$ ) and *H. influenzae* (MIC range 0.8-1.6  $\mu\text{g/ml}$ ). However, BC-3311 had no relevant activity against *E. coli* (MIC



6.4 µg/ml). Compared to BC-3311, pleuromutilin consisting of only the tricyclic pleuromutilin core and BC-3291 being characterized by a rather bulky extended methyl-piperazine-side-chain showed an approximately 16- to 32-fold lower activity (Table 3).

**Chemical footprinting of pleuromutilin derivatives with *E. coli* ribosomes.** From previous studies performed with *E. coli* ribosomes, it was shown that pleuromutilins hinder peptide bond formation by binding to the PTC of the large ribosomal subunit. To determine whether the improved activities of BC-3311 can be attributed to additional or different interaction sites at the bacterial ribosomal target (23S rRNA), chemical probing followed by primer extension analysis was performed. Briefly, 70S ribosome-antibiotic complexes as well as control samples were subjected to modification with DMS and CMCT. Alterations in the chemical reactivity of the 23S rRNA target nucleotides induced by binding of the antibiotic to the ribosome were monitored by reverse transcription/primer extension reactions using primers designed to cover hairpin 35 of domain II and the entire domain V of *E. coli* 23S rRNA (Table 1). The native pleuromutilin core and BC-3291 with its bulky extended methyl-piperazine-side-chain were included as non-active control antibiotics in the study. The derivatives tiamulin, valnemulin, azamulin and the macrolide erythromycin were included as active reference antibiotics. The observed protection and enhancement patterns of chemical reactivity by the antibiotics are summarized in Table 4 and indicated on the secondary structure of the PTC region of the 23S rRNA (Fig. 2A). Footprints were initially analyzed over a range of different concentrations of antibiotics for the same amount of ribosomes to identify the optimal concentrations to use for the present study (data not shown). Chemical footprinting analysis with BC-3311, BC-3291 and the other pleuromutilin derivatives showed enhancement at nucleotides A2058 and A2059, protection at nucleotides U2584 and U2585 and strong protection at nucleotide U2506, as described earlier (data not shown). The control erythromycin showed footprints as described previously in the literature : protections at A752 (domain II), A2058 and A2059 (domain V). For all tested pleuromutilins, no footprinting was observed within helix 35 of domain II of the 23S rRNA. This is not unexpected since, although some macrolide antibiotics interact with domain II, this region is located adjacent to the PTC and does not comprise the pleuromutilin binding site.

**Chemical footprinting of pleuromutilin derivatives with *S. aureus* ribosomes.** Because pleuromutilins are highly active against gram-positive organisms and are not active against the gram-negative *E. coli*, we were interested in examining the question if the interaction sites of pleuromutilin derivatives with the *S. aureus* 23S rRNA differ from those determined with the *E. coli* 23S rRNA. Since the 23S rRNA nucleotide sequence of *S. aureus* varies from that of *E. coli*, new primers were designed for the reverse transcription of domain V of *S. aureus* 23S rRNA (Table 1). The primers were chosen to anneal at almost identical positions as with the *E. coli* 23S rRNA. Only the position of primer P092-2Sa was shifted upstream to a less structured region to achieve better annealing to the 23S rRNA target. All observed protection and enhancement effects are tabulated in Table 4 and indicated on the secondary structure of domain V of *S. aureus* 23S rRNA (Fig. 2B). As described for the *E. coli* 23S rRNA, analysis of the footprints with the novel pleuromutilin derivatives BC-3311 and BC-3291 showed enhancements at nucleotides A2058 and A2059 (*E. coli* numbering, data not shown) and protections at nucleotides U2584 and U2585 (*E. coli* numbering, Fig. 3A). In addition, protection at nucleotide U2506 (*E. coli* numbering), which was observed with the *E. coli* ribosome, could not be detected with the *S. aureus* ribosome (Fig. 3B). Further, novel interaction sites with the 23S rRNA of *S. aureus* were revealed, which were not detected with *E. coli*: nucleotides U2533 and U2563 (*E. coli* numbering) were protected (Fig. 3A) and the accessibility of nucleotide 2192 (corresponding to C2219 in *S. aureus* and U2192 in *E. coli*) for chemical modification was enhanced by BC-3311, BC-3291 and the other pleuromutilin derivatives (Fig. 3C, 3D). No significant differences in the intensities of footprints could be detected when comparing the various pleuromutilin derivatives. In summary, the chemical probing data indicated that pleuromutilin derivatives interact differently with the 23S rRNA domain V of *S. aureus* than with the 23S rRNA domain V of *E. coli*. The novel pleuromutilin derivatives BC-3311 and BC-3291 showed overlapping interaction sites with other pleuromutilins.

## DISCUSSION

Here, we report the structure of two novel pleuromutilin derivatives, BC-3311 and BC-3291, their *in vitro* antibacterial activities against representative bacterial species and the analysis of their interaction sites with the ribosomal targets of the model organism *E. coli* and the target pathogen *S. aureus*.

The novel semi-synthetic pleuromutilin derivative BC-3311 with an azepan-4-ylsulfanyl-side-chain displayed potent antibacterial activity against the major pathogens involved in RTIs (e.g. *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*) and SSSIs (e.g. *S. aureus* including MRSA, *S. pyogenes* and other coagulase negative staphylococci) (Tables 2, 3). In contrast, BC-3291 with an extended methyl-piperazine-side-chain, exhibited significantly lower antibacterial activity, which was comparable to that of the pleuromutilin core without a C14-extension.

Differences in antibacterial activities of pleuromutilin derivatives against the same bacterial pathogen have been attributed to the chemical nature of their respective side-chain. While the tricyclic pleuromutilin core binds the PTC, the side-chain-extensions of the derivatives are seen to adopt different conformations within the cavity thus stabilizing the pleuromutilin-ribosome complex. Maximizing the number of hydrogen bonds, e.g. by lengthening of the C14-side-chain, has been considered as a providing potential to further improve the antibacterial activity of the pleuromutilins. Therefore, BC-3291 was designed with an extended methyl-piperazine-side-chain to maximize the number of hydrogen bond interactions. Surprisingly, only BC-3311 with a compact and flexible cyclic azepan-4-ylsulfanyl-extension exhibited potent antibacterial activity, whereas BC-3291 was inactive. To understand how the sidechains may influence binding, we have generated models of BC-3291 and BC-3311 (as well as azamulin and valnemulin) bound to the ribosome (Fig. 4) based on the known crystal structures of pleuromutilins in complex with the *D. radiodurans* 50S subunit. BC-3311 binds at the PTC, overlapping aminoacyl moieties of both A- and P-site bound tRNAs (Fig. 4A, 4B). The longer side-chains of BC-3291 and valnemulin (Fig. 4C) would be expected to penetrate deeper into the space between nucleotides C2063 and U2586 (Fig. 4D, 4E), as seen to a lesser extent for tiamulin (Fig. 4F).

To determine the interaction sites of the novel pleuromutilin derivatives with the ribosomal target, chemical footprinting of the 23S rRNA domain V of *E. coli* and *S. aureus* was performed; also included in the study were the pleuromutilin

core as well as tiamulin, azamulin and valnemulin. *S. aureus* was used since it is the target pathogen of the derivatives and no analysis of interaction sites of pleuromutilins with ribosomes of gram-positive bacteria have been reported thus far. Analysis of the footprints revealed differences of pleuromutilin-ribosome interaction sites in *S. aureus* compared with *E. coli*. Beside the enhancements at nucleotides A2058, A2059 and protections at nucleotides U2584, U2585, which were also observed in *E. coli*, the following differences in modifications were found: no protection of U2506, located at the heart of the PTC, an enhancement at nucleotide 2192 (C2219 in *S. aureus* and U2192 in *E. coli*) and protection at U2533 and U2563 absent in *E. coli*. Such species-specific interactions with the 23S rRNA target have also been described for *B. hyodysenteriae*. Compared to *E. coli*, only U2506 was protected by tiamulin, whereas no modifications were found at U2584 and U2585 of the 23S rRNA of *B. hyodysenteriae*. The PTC is highly conserved and contains many universally conserved nucleotides, therefore one might expect that antibiotics would interact with ribosomes from different species in an identical manner. However, comparisons of binding of identical antibiotics to different species of ribosomes reveals that small differences in sequence in less conserved regions can influence the binding of drug directly, but also indirectly by perturbing the conformation of the neighbouring conserved residues. Indeed, resistance to pleuromutilins generally results from alterations of non-conserved nucleotides residing in the close vicinity of the PTC (but not directly within the conserved binding site) and/or mutations within the loop of ribosomal protein L3 that extends towards the PTC.

Mapping the footprinted nucleotides onto the X-ray structure of *D. radiodurans* 50S subunit reveals the location of the nucleotides relative to the pleuromutilin binding site (Fig. 4A, 4D, 4E). The enhancements of A2058 and A2059 are likely to be indirect effects resulting from conformational changes since these residues are located 7-8 Å from the pleuromutilin core towards the peptide exit tunnel. Similarly, the enhancement of C2192 (*E. coli* numbering, corresponding to C2219 in *S. aureus*) on the *S. aureus* ribosomes appears to be indirect as C2192 is located in helix 76 on the L1 stalk (Fig. 4A). The L1 stalk is very flexible and has been postulated to oscillate between different positions during the elongation cycle. Whether antibiotic binding induces a specific conformation of the stalk is intriguing but remains to be clarified. The two novel protections at U2533 and U2563 in *S. aureus* and absent in *E. coli* are close

together (with  $\sim 18$  Å): U2533 is located in helix H91, whereas U2563 is located at the three-way junction between H91, H92 and H93. However, these nucleotides are located far from the PTC (50-70 Å away) and L3 (25 Å), suggesting the possibility that a second pleuromutilin molecule interacts with the ribosome, as seen for the tetracyclines and kasugamycin. Curiously, this region comprises part of the binding pocket for translation factors, however further work will be needed to confirm any effect of pleuromutilins on translation factor activity.

One unexpected observation from our study is that no significant differences in footprint intensities were observed when comparing the six different pleuromutilin derivatives. This is especially surprising given the distinct structures and activities that these compounds display, as well as the different binding modes that are observed for various pleuromutilin side-chains from the available X-ray structures (Fig. 4). Clearly, one explanation for this is that the additional side-chain-specific interactions that could differ between the derivatives may not be detectable with the chemical probing method since this technique can only provide information about nucleotides that are accessible to the chemical probes. Despite this limitation, we were able to identify three nucleotides, U2506, U2584 and U2585, that are protected upon binding of all pleuromutilins to *E. coli* ribosomes, consistent with the fact that these nucleotides form part of the pleuromutilin binding pocket (Fig. 4D-F). Indeed, binding of pleuromutilins appears to utilize an induced-fit mechanism, involving conformational changes in U2506 and U2585 that establish interaction with the pleuromutilin core and side-chain, respectively. Based on these X-ray structures, it would appear that the protection of U2585 and U2506 may result from interaction between these nucleotides, rather than with the drug (Fig. 4D). However, these protections are also observed with the pleuromutilin core that lacks any C14-side-chain, suggesting that the pleuromutilin core also induces a similar conformational change in the PTC. Surprisingly, this appears not to be the case in *S. aureus* where no protection of U2506 was observed (Fig. 3B), implying that the pleuromutilin binding pocket differs from that of *E. coli* despite the high conservation between these two species. This finding reinforces the idea that sequence variation within non-conserved regions of the PTC of different species leads to conformational differences within the drug binding site, emphasizing the need to study the binding and action of drugs on their target species.

Taken together, the novel pleuromutilin derivative BC-3311 displays very high antibacterial activities against clinical isolates associated with RTIs and SSSIs including antibiotic resistant staphylococcal and streptococcal strains. Chemical footprinting analysis demonstrated novel interaction sites of BC-3311 and other pleuromutilins with the 23S rRNA ribosomal target of *S. aureus* compared to that of *E. coli*. However, further work will be necessary to understand the exact molecular basis of the high activity of novel pleuromutilin derivatives against sensitive and resistant clinical isolates. This implies obtaining better insight into the interaction of the derivatives with their gram-positive ribosomal targets and to elucidate their exact inhibitory mode of action during protein synthesis. This knowledge will be beneficial for future rational design of novel compounds efficacious against bacterial pathogens resistant to current antibiotics.

## ACKNOWLEDGMENTS

This work was supported by FFG (Österreichische Forschungsförderungsgesellschaft mbH, grant no. 812138 to E.C.), ZIT (Zentrum für Innovation und Technologie GmbH, Vienna 2006, grant no. 205509 to Nabriva) and DFG (Deutsche Forschungsgemeinschaft, grant no. WI3285/1-1 to D.N.W.). We acknowledge I. Chopra, F.J. Schmitz, the General Hospital of Vienna, Novartis and the Clinical Microbiology Department of the University College London Hospitals for supplying clinical isolates. We thank Werner Heilmayer and Rosemarie Mang for continuing support in chemical questions. We also thank Rickmer Dey, Attila Hegyi and Astrid Gruss for technical support in microbiological experiments and ribosome preparations, and Karine Gonzales and Christina Waldsich for their advice with chemical footprinting experiments.

---

**REFERENCES**

1. **Baquero, F.** 1997. Gram-positive resistance: challenge for the development of new antibiotics. *J. Antimicrob. Chemother.* **39 Suppl A**:1-6.
2. **Bartlett, J.** 2000. Treatment of community-acquired pneumonia. *Chemotherapy* **46 Suppl 1**:24-31.
3. **Brodersen, D. E., W. M. Clemons, Jr., A. P. Carter, R. J. Morgan-Warren, B. T. Wimberly, and V. Ramakrishnan.** 2000. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* **103**:1143-1154.
4. **Brooks, G., W. Burgess, D. Colthurst, J. D. Hinks, E. Hunt, M. J. Pearson, B. Shea, A. K. Takle, J. M. Wilson, and G. Woodnutt.** 2001. Pleuromutilins. Part 1. The identification of novel mutilin 14-carbamates. *Bioorg. Med. Chem.* **9**:1221-1231.
5. **Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren.** 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**:905-921.
6. **Cannone, J. J., S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Muller, N. Pande, Z. Shang, N. Yu, and R. R. Gutell.** 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**:1-31.
7. **Champney, W. S., and W. K. Rodgers.** 2007. Retapamulin inhibition of translation and 50S ribosomal subunit formation in *Staphylococcus aureus* cells. *Antimicrob. Agents Chemother.* **51**:3385-3387.
8. **CLSI (Clinical and Laboratory Standards Institute (formerly NCCLS)).** 2006. Methods for dilution antimicrobial susceptibility test for bacteria that grow aerobically; approved standard, 7th ed. Clinical and Laboratory Standards Institute document M7-A7. Vol. 26 No. 2. Clinical and Laboratory Standards Institute, Wayne, PA. ed.
9. **CLSI (Clinical and Laboratory Standards Institute (formerly NCCLS)).** 2006. Performance standards for antimicrobial susceptibility testing; 16th



- informational supplement. Clinical and Laboratory Standards Institute document M100-S16. Vol. 26 No. 3. Clinical and Laboratory Standards Institute, Wayne, PA. ed.
10. **Davidovich, C., A. Bashan, T. Auerbach-Nevo, R. D. Yaggie, R. R. Gontarek, and A. Yonath.** 2007. Induced-fit tightens pleuromutilins binding to ribosomes and remote interactions enable their selectivity. *Proc. Natl. Acad. Sci. U. S. A.* **104**:4291-4296.
  11. **Dunbar, L. M.** 2003. Current issues in the management of bacterial respiratory tract disease: the challenge of antibacterial resistance. *Am. J. Med. Sci.* **326**:360-368.
  12. **Egger, H., and H. Reinshagen.** 1976. New pleuromutilin derivatives with enhanced antimicrobial activity. I. Synthesis. *J. Antibiot. (Tokyo)* **29**:915-922.
  13. **Egger, H., and H. Reinshagen.** 1976. New pleuromutilin derivatives with enhanced antimicrobial activity.II.Structure-activity correlations. *J. Antibiot. (Tokyo)* **29**:923-927.
  14. **Ehresmann, C., F. Baudin, M. Mougel, P. Romby, J. P. Ebel, and B. Ehresmann.** 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**:9109-9128.
  15. **Emsley, P., and K. Cowtan.** 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**:2126-2132.
  16. **Hansen, L. H., P. Mauvais, and S. Douthwaite.** 1999. The macrolide-ketolide antibiotic binding site is formed by structures in domains II and V of 23S ribosomal RNA. *Mol. Microbiol.* **31**:623-631.
  17. **Harms, J. M., D. N. Wilson, F. Schluenzen, S. R. Connell, T. Stachelhaus, Z. Zaborowska, C. M. Spahn, and P. Fucini.** 2008. Translational regulation via L11: molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. *Mol. Cell* **30**:26-38.
  18. **Heilmann, C., L. Jensen, J. S. Jensen, K. Lundstrom, D. Windsor, H. Windsor, and D. Webster.** 2001. Treatment of resistant mycoplasma infection in immunocompromised patients with a new pleuromutilin antibiotic. *J. Infect.* **43**:234-238.
  19. **Hodgin, L. A., and G. Hogenauer.** 1974. The mode of action of pleuromutilin derivatives. Effect on cell-free polypeptide synthesis. *Eur. J. Biochem.* **47**:527-533.

20. **Hogenauer, G.** 1975. The mode of action of pleuromutilin derivatives. Location and properties of the pleuromutilin binding site on *Escherichia coli* ribosomes. *Eur. J. Biochem.* **52**:93-98.
21. **Jacobs, M. R., J. Anon, and P. C. Appelbaum.** 2004. Mechanisms of resistance among respiratory tract pathogens. *Clin. Lab. Med.* **24**:419-453.
22. **Jones, R. N., T. R. Fritsche, H. S. Sader, and J. E. Ross.** 2006. Activity of retapamulin (SB-275833), a novel pleuromutilin, against selected resistant gram-positive cocci. *Antimicrob. Agents Chemother.* **50**:2583-2586.
23. **Kavanagh, F., A. Hervey, and W. J. Robbins.** 1951. Antibiotic substances from basidiomycetes: VIII. *Pleurotus Multilus (Fr.) Sacc.* and *Pleurotus Passeckerianus Pilat.* *Proc. Natl. Acad. Sci. U. S. A.* **37**:570-574.
24. **Kosowska-Shick, K., C. Clark, K. Credito, P. McGhee, B. Dewasse, T. Bogdanovich, and P. C. Appelbaum.** 2006. Single- and multistep resistance selection studies on the activity of retapamulin compared to other agents against *Staphylococcus aureus* and *Streptococcus pyogenes*. *Antimicrob. Agents Chemother.* **50**:765-769.
25. **Long, K. S., L. H. Hansen, L. Jakobsen, and B. Vester.** 2006. Interaction of pleuromutilin derivatives with the ribosomal peptidyl transferase center. *Antimicrob. Agents Chemother.* **50**:1458-1462.
26. **Malbruny, B., A. Canu, B. Bozdogan, B. Fantin, V. Zarrouk, S. Dutka-Malen, C. Feger, and R. Leclercq.** 2002. Resistance to quinupristin-dalfopristin due to mutation of L22 ribosomal protein in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **46**:2200-2207.
27. **Nichols, R. L.** 1999. Optimal treatment of complicated skin and skin structure infections. *J. Antimicrob. Chemother.* **44 Suppl A**:19-23.
28. **Nicolau, D.** 2002. Clinical and economic implications of antimicrobial resistance for the management of community-acquired respiratory tract infections. *J. Antimicrob. Chemother.* **50 Suppl S1**:61-70.
29. **Pankuch, G. A., G. Lin, D. B. Hoellman, C. E. Good, M. R. Jacobs, and P. C. Appelbaum.** 2006. Activity of retapamulin against *Streptococcus pyogenes* and *Staphylococcus aureus* evaluated by agar dilution, microdilution, E-test, and disk diffusion methodologies. *Antimicrob. Agents Chemother.* **50**:1727-1730.
30. **Pioletti, M., F. Schlunzen, J. Harms, R. Zarivach, M. Gluhmann, H. Avila, A. Bashan, H. Bartels, T. Auerbach, C. Jacobi, T. Hartsch, A.**

- Yonath, and F. Franceschi.** 2001. Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J.* **20**:1829-1839.
31. **Poulsen, S. M., M. Karlsson, L. B. Johansson, and B. Vester.** 2001. The pleuromutilin drugs tiamulin and valnemulin bind to the RNA at the peptidyl transferase centre on the ribosome. *Mol. Microbiol.* **41**:1091-1099.
32. **Poulsen, S. M., C. Kofoed, and B. Vester.** 2000. Inhibition of the ribosomal peptidyl transferase reaction by the mycarose moiety of the antibiotics carbomycin, spiramycin and tylosin. *J. Mol. Biol.* **304**:471-481.
33. **Pringle, M., J. Poehlsgaard, B. Vester, and K. S. Long.** 2004. Mutations in ribosomal protein L3 and 23S ribosomal RNA at the peptidyl transferase centre are associated with reduced susceptibility to tiamulin in *Brachyspira* spp. isolates. *Mol. Microbiol.* **54**:1295-1306.
34. **Rittenhouse, S., S. Biswas, J. Broskey, L. McCloskey, T. Moore, S. Vasey, J. West, M. Zalacain, R. Zonis, and D. Payne.** 2006. Selection of retapamulin, a novel pleuromutilin for topical use. *Antimicrob. Agents Chemother.* **50**:3882-3885.
35. **Rittenhouse, S., C. Singley, J. Hoover, R. Page, and D. Payne.** 2006. Use of the surgical wound infection model to determine the efficacious dosing regimen of retapamulin, a novel topical antibiotic. *Antimicrob. Agents Chemother.* **50**:3886-3888.
36. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY ed. Cold Spring Harbor Laboratory Press.
37. **Schlunzen, F., C. Takemoto, D. N. Wilson, T. Kaminishi, J. M. Harms, K. Hanawa-Suetsugu, W. Szaflarski, M. Kawazoe, M. Shirouzu, K. H. Nierhaus, S. Yokoyama, and P. Fucini.** 2006. The antibiotic kasugamycin mimics mRNA nucleotides to destabilize tRNA binding and inhibit canonical translation initiation. *Nat. Struct. Mol. Biol.* **13**:871-878.
38. **Schlunzen, F., E. Pyetan, P. Fucini, A. Yonath, and J. M. Harms.** 2004. Inhibition of peptide bond formation by pleuromutilins: the structure of the 50S ribosomal subunit from *Deinococcus radiodurans* in complex with tiamulin. *Mol. Microbiol.* **54**:1287-1294.

39. **Selmer, M., C. M. Dunham, F. V. t. Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir, and V. Ramakrishnan.** 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**:1935-1942.
40. **Stratton, C. W.** 2004. Antimicrobial resistance in respiratory tract pathogens. *Expert Rev. Anti. Infect. Ther.* **2**:641-647.
41. **Stulberg, D. L., M. A. Penrod, and R. A. Blatny.** 2002. Common bacterial skin infections. *Am. Fam. Physician.* **66**:119-124.
42. **Swaney, S. M., H. Aoki, M. C. Ganoza, and D. L. Shinabarger.** 1998. The oxazolidinone linezolid inhibits initiation of protein synthesis in bacteria. *Antimicrob. Agents Chemother.* **42**:3251-3255.
43. **Thompson, J., W. E. Tapprich, C. Munger, and A. E. Dahlberg.** 2001. *Staphylococcus aureus* domain V functions in *Escherichia coli* ribosomes provided a conserved interaction with domain IV is restored. *RNA* **7**:1076-1083.
44. **Wilson, D. N., J. M. Harms, K. H. Nierhaus, F. Schlunzen, and P. Fucini.** 2005. Species-specific antibiotic-ribosome interactions: implications for drug development. *Biol. Chem.* **386**:1239-1252.
45. **Wilson, D. N., and K. H. Nierhaus.** 2005. Ribosomal proteins in the spotlight. *Crit. Rev. Biochem. Mol. Biol.* **40**:243-267.
46. **Yan, K., L. Madden, A. E. Choudhry, C. S. Voigt, R. A. Copeland, and R. R. Gontarek.** 2006. Biochemical characterization of the interactions of the novel pleuromutilin derivative retapamulin with bacterial ribosomes. *Antimicrob. Agents Chemother.* **50**:3875-3881.

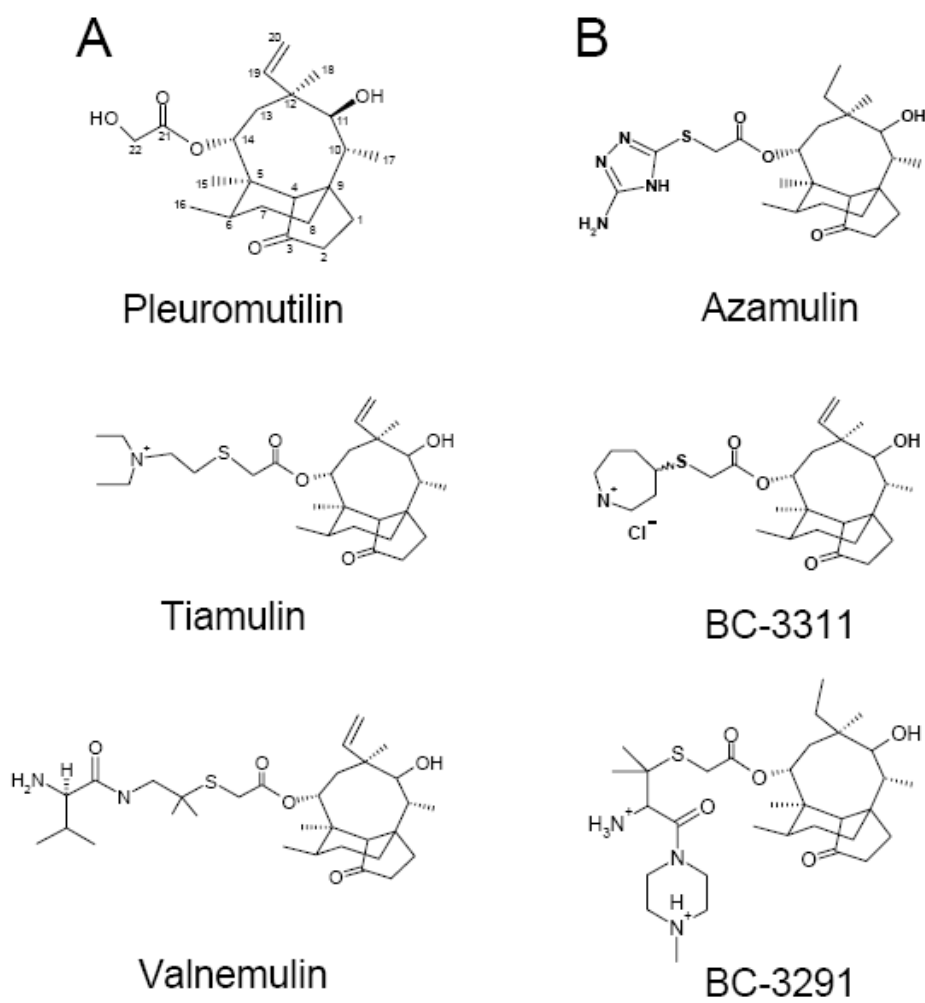
## FIGURE LEGENDS

**Fig. 1.** Chemical structures of pleuromutilin and C14-sulfanyl acetate derivatives used in this study. (A) Pleuromutilin and derivatives with linear side-chain at position C14 on the octane ring. (B) Derivatives with cyclic side-chain at position C14 on the octane ring.

**Fig. 2.** Secondary structures of the PTC and adjacent regions of domain V of *E. coli* (A) and *S. aureus* (B) 23S rRNA showing the chemical footprints of BC-3311, BC-3291, pleuromutilin, other conventional pleuromutilin derivatives and erythromycin. The secondary structures are modifications of structure models obtained from the Comparative RNA Web Site at URL: <http://www.rna.cccb.utexas.edu/>. The nucleotide positions exhibiting altered reactivity in the presence of antibiotics (Aza, Azamulin; Ery, Erythromycin; Ple, Pleuromutilin; Tia, Tiamulin; Val, Valnemulin; BC-3291 and BC-3311) are indicated. Footprints obtained in this study are shown in red (DMS) and green (CMCT). Enclosed circles represent protections and open circles denote enhancements.

**Fig. 3.** Gel autoradiograms showing chemical footprints of antibiotics in the central loop region of domain V of *S. aureus* (A, B, C) and *E. coli* (D) 23S rRNA. Ribosomes were modified in the presence of the following antibiotics: Azamulin (Aza), Erythromycin (Ery), Pleuromutilin (Ple), Tiamulin (Tia), Valnemulin (Val), BC-3291 and BC-3311. Wedges indicate the increase in antibiotic concentrations: 100, 250, 500  $\mu$ M (corresponding to ribosome:antibiotic ratios of 1:5, 1:12.5 and 1:25). Nucleotides displaying altered reactivity in the presence of antibiotics are indicated (P, protection; E, enhancement). (A) CMCT modifications detected with primer P095Sa. (B) CMCT modifications detected by reverse transcription using primer P094Sa. (C) DMS modifications detected with primer P093Sa. (D) CMCT modifications detected with primer P093. A, C, G and U lanes designate dideoxy sequencing reactions. Lane "NM" indicates reactions with chemically unmodified 70S ribosomes in the absence of antibiotic. Lane "ND" indicates reactions with chemically modified 70S ribosomes in the absence of antibiotic. The reverse transcriptase stops one nucleotide before the corresponding nucleotide in the sequencing tracks (Table 3).

**Fig. 4.** The pleuromutilin binding site on the large ribosomal subunit. (A) Overview of the pleuromutilin binding site (green) in the PTC relative to P-tRNA (blue; PDB2J00; ) and a modeled polypeptide chain in the exit tunnel (tan). The rRNA and r-proteins (PDB2ZJR; ) are coloured grey, except for nucleotides U2192 (red), U2533 (orange) and U2563 (cyan). (B) Relative position of the pleuromutilins, such as BC-3311 (green), compared to A- (blue) and P-tRNA (yellow). Selected nucleotides of the PTC are coloured green and shown as a surface representation. (C) Comparison of bound positions of tiamulin (orange; PDB1XBP; ) with models for BC-3311 (green), BC-3291 (pink), valnemulin (pale blue) and azamulin (yellow). (D)-(F) Models for binding pockets of (D) BC-3311, (E) BC-3291 compared with that of (F) tiamulin. Hydrogen bonds are represented by dashed lines. The arrow in (D) indicates the conformational change observed in U2506 upon binding of pleuromutilins to the PTC , whereas the arrow in (E) indicates the conformational flexibility of the extended C14-side-chain of BC-3291. All figures were made with pymol (<http://www.pymol.org>).

**Fig. 1**

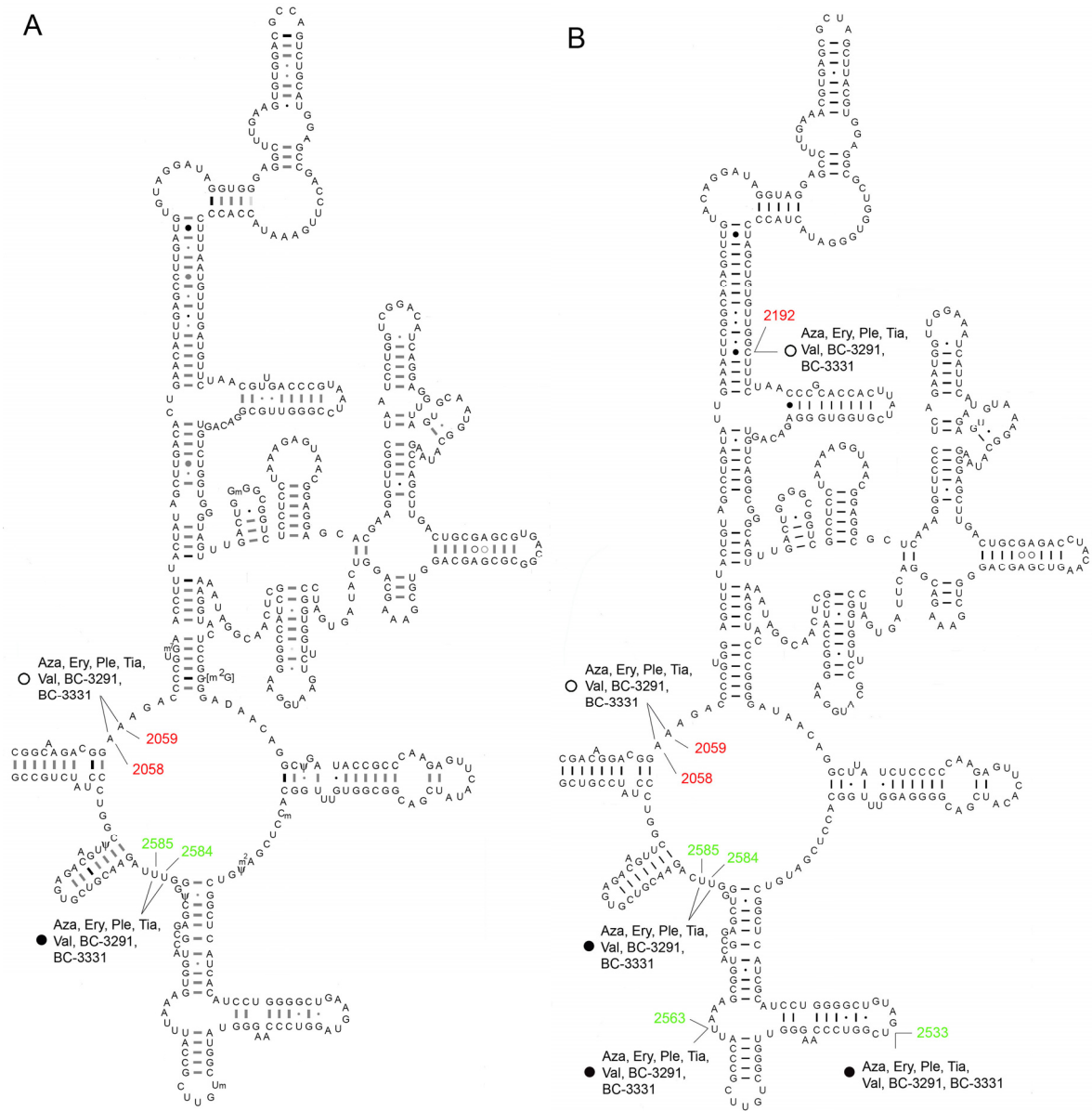


Fig. 2



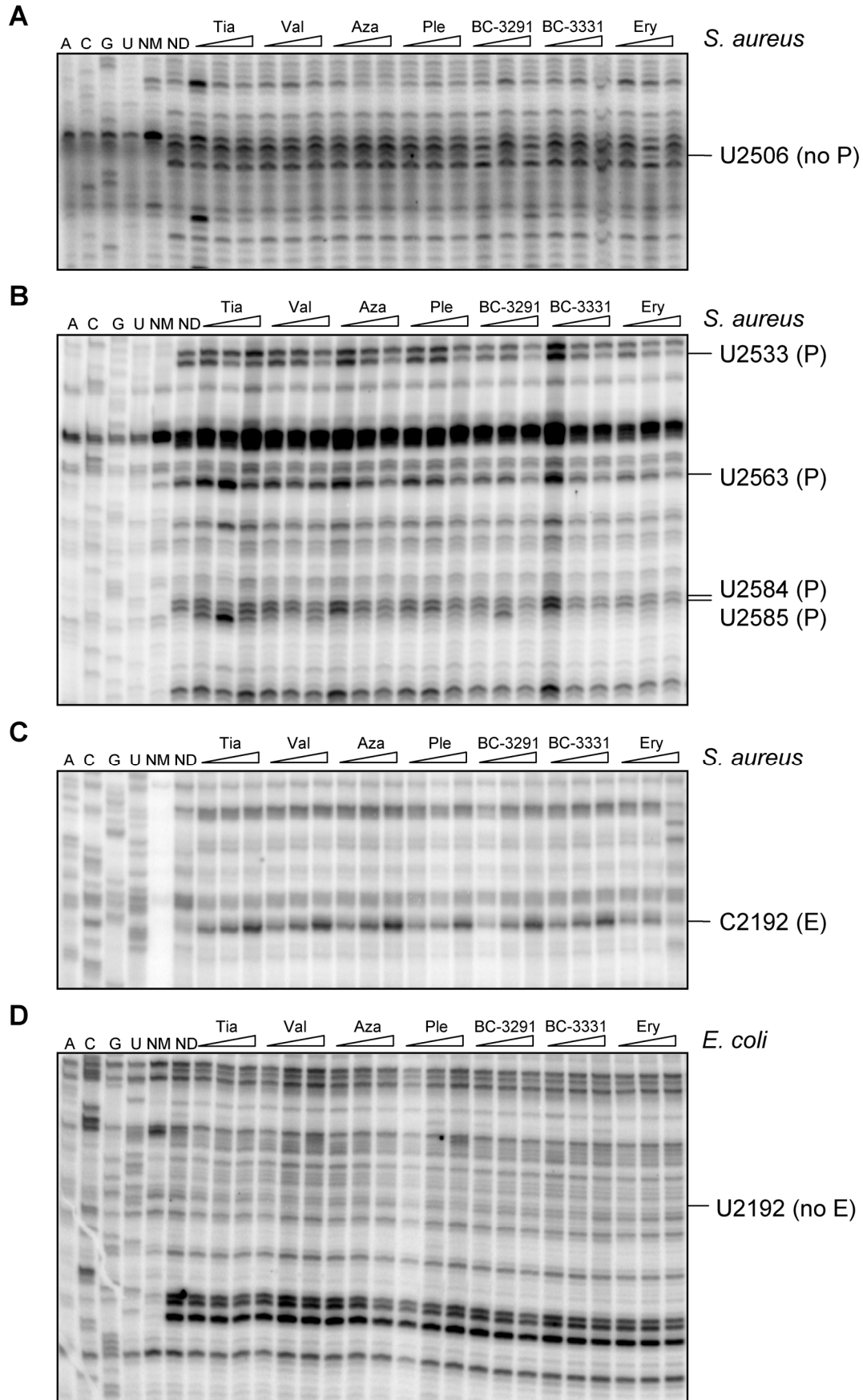


Fig. 3

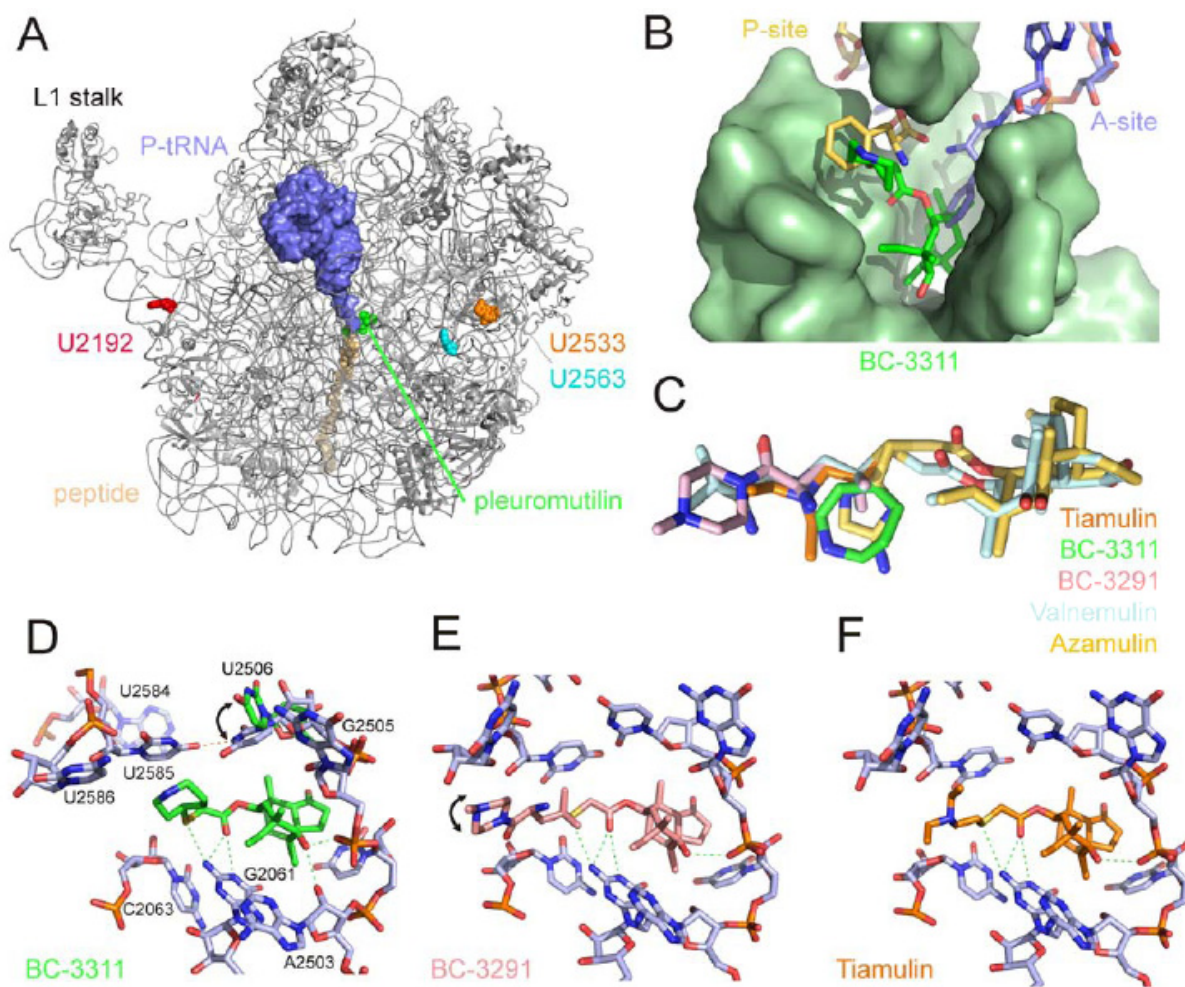


Fig. 4

TABLE 1. Primers used for chemical footprinting and sequencing reactions

Primer	Sequence (5'-3')	Complementary to nucleotides <sup>a</sup>	23 S rRNA domain	Reference
<i>E. coli</i>				
P091	GGCGCTACCTAAATAGCT	821-838	II, hairpin 35	(31)
P092	CCATGCAGACTGGCGTC	2141-2157	V	(31)
P093	CCGACCAGGATTAGCCAACC	2289-2308	V	(31)
P094	TCGCGTACCACTTTA	2563-2577	V	(31)
P095	TCCGTTCTCTCGTACT	2654-2670	V	(31)
<i>S. aureus</i>				
P091Sa	GGCTAGCCCTAAAGCTATTTTC	818-838 (863-883)	II, hairpin 35	This study
P092-2Sa	TTCAAAGGCTCCTACCTATCCTGTAC	2110-2135 (2137-2162)	V	This study
P093Sa	CCAACCATTCTGAGGGAACC	2289-2308 (2316-2335)	V	This study
P094Sa	GCTCGCGTACCGCTTTAATG	2560-2579 (2587-2606)	V	This study
P095Sa	GATCCCGGTCCTCTCGTAC	2655-2673 (2682-2700)	V	This study

<sup>a</sup> Nucleotides are numbered according to the sequence of *E. coli*. *S. aureus* numberings are in brackets.

TABLE 2. Antibacterial activity of BC-3291, BC3311, conventional pleuromutilin derivatives and standard antibiotics against target pathogens involved in RTIs and SSSIs and other infections

Species	Strain	Source	MIC ( $\mu\text{g/ml}$ )									
			BC-3311	BC-3291	PLE	TIA	AZA	VAL	LZD	AZM	MXF	DOX
<i>S. aureus</i> (MSSA)	B6	ATCC 10390	0.1	1.6	1.6	0.4	0.8	0.05	3.2	0.8	0.05	0.8
<i>S. aureus</i> (MSSA)	B7	ATCC 29213	0.1	1.6	1.6	0.4	0.8	0.05	3.2	0.8	0.2	0.8
<i>S. aureus</i> (MSSA)	B9	ATCC 49951	0.1	1.6	0.8	0.4	0.4	0.05	3.2	0.8	0.05	0.4
<i>S. aureus</i> (MSSA)	B302	Clin. isolate <sup>a</sup>	0.05	1.6	0.8	0.2	0.4	0.05	1.6	0.4	0.4	0.4
<i>S. aureus</i> (MRSA)	B29	Clin. isolate <sup>a</sup>	0.05	1.6	0.8	0.4	0.8	0.1	1.6	>25.6	1.6	12.8
<i>S. aureus</i> (MRSA)	$\Delta$ 2468	Clin. isolate <sup>c</sup>	0.1	1.6	1.6	0.4	0.8	0.1	3.2	>25.6	0.4	25.6
<i>S. aureus</i> (MRSA)	B252	Clin. isolate <sup>a</sup>	0.1	1.6	1.6	0.4	0.8	0.05	3.2	25.6	0.05	0.4
<i>S. aureus</i> (MRSA)	B246	Clin. isolate <sup>a</sup>	0.1	1.6	1.6	0.4	0.4	0.05	3.2	1.6	0.05	0.4
<i>S. epidermidis</i> (MRSE)	B 534	Clin. isolate <sup>b</sup>	0.1	1.6	1.6	0.2	0.4	0.1	1.6	>25.6	3.2	1.6
<i>S. epidermidis</i>	B 573	Clin. isolate <sup>b</sup>	0.05	1.6	1.6	0.2	0.4	0.1	3.2	>25.6	0.1	0.4
<i>S. epidermidis</i>	$\Delta$ 1570	Clin. isolate <sup>c</sup>	0.1	3.2	1.6	0.4	0.4	0.05	3.2	1.6	0.2	0.4
<i>S. haemolyticus</i>	B549	Clin. isolate <sup>b</sup>	0.025	0.8	0.4	0.2	0.4	0.05	0.8	>25.6	1.6	1.6
<i>S. haemolyticus</i>	B565	Clin. isolate <sup>b</sup>	0.05	1.6	0.8	0.2	0.8	0.05	0.8	>25.6	1.6	1.6
<i>S. pyogenes</i>	B780	Clin. isolate <sup>b</sup>	0.02	1.28	1.28	0.16	0.32	0.01	2.56	0.02	1.28	>2.56
<i>S. pyogenes</i>	B781	Clin. isolate <sup>b</sup>	0.02	1.28	2.56	0.16	0.32	0.01	>2.56	0.02	0.32	0.32
<i>S. pyogenes</i>	B782	Clin. isolate <sup>b</sup>	0.005	0.32	1.28	0.08	0.16	$\leq$ 0.0025	1.28	0.04	0.64	>2.56
<i>S. pyogenes</i>	B784	Clin. isolate <sup>b</sup>	0.005	0.32	0.64	0.04	0.16	$\leq$ 0.0025	1.28	0.01	0.64	>2.56
<i>E. faecium</i>	$\Delta$ 459	Clin. isolate <sup>e</sup>	0.05	1.6	1.6	0.4	0.4	0.05	1.6	>25.6	6.4	>25.6
<i>E. faecium</i> (VRE)	B222	Clin. isolate <sup>a</sup>	0.05	1.6	1.6	0.4	0.4	0.05	1.6	>25.6	6.4	0.2
<i>E. faecium</i> (VRE)	B207	Clin. isolate <sup>a</sup>	0.05	1.6	0.8	0.4	0.4	0.05	1.6	>25.6	3.2	25.6
<i>E. faecium</i>	B215	Clin. isolate <sup>a</sup>	0.2	6.4	3.2	0.8	0.8	0.1	1.6	3.2	>25.6	0.4
<i>S. pneumoniae</i>	B11	ATCC 49619	0.16	0.64	3.2	0.32	0.64	0.08	1.28	0.08	0.16	0.16
<i>S. pneumoniae</i>	B415	Clin. isolate <sup>c</sup>	0.08	0.64	3.2	0.4	1.28	0.16	0.64	0.04	0.16	0.08
<i>S. pneumoniae</i>	B37	Clin. isolate <sup>e</sup>	0.16	1.28	3.2	3.2	0.64	0.08	1.28	0.04	0.16	0.16
<i>S. pneumoniae</i>	L6/3	Clin. isolate <sup>d</sup>	0.16	0.64	6.4	0.2	1.28	0.08	0.64	0.64	0.16	0.16
<i>S. pneumoniae</i>	L4	Clin. isolate <sup>d</sup>	0.08	0.64	3.2	0.4	1.28	0.08	1.28	0.04	0.16	>2.56
<i>S. pneumoniae</i>	B78	Clin. isolate <sup>a</sup>	0.08	1.28	3.2	0.2	1.28	0.08	0.64	0.04	0.32	>2.56
<i>S. pneumoniae</i>	B225	Clin. isolate <sup>a</sup>	0.16	0.64	12.8	0.4	2.56	0.08	1.28	>2.56	0.16	2.56
<i>H. influenzae</i>	B35	ATCC 49247	1.6	>25.6	6.4	6.4	3.2	6.4	12.8	1.6	0.01	0.2
<i>H. influenzae</i>	B36	ATCC 49766	1.6	>25.6	12.8	12.8	6.4	6.4	12.8	3.2	0.02	0.4
<i>H. influenzae</i>	B402	ATCC 43335	0.8	>25.6	6.4	6.4	3.2	3.2	25.6	3.2	0.08	0.4
<i>H. influenzae</i>	B405	ATCC 43334	0.8	25.6	3.2	6.4	1.6	1.6	12.8	1.6	0.08	0.8
<i>M. catarrhalis</i>	B22	Clin. isolate <sup>c</sup>	0.025	0.8	0.2	0.4	0.4	0.05	6.4	0.025	0.1	0.8
<i>M. catarrhalis</i>	B407	ATCC 43618	0.025	0.4	0.2	0.1	0.2	$\leq$ 0.0125	3.2	$\leq$ 0.0125	0.1	0.8
<i>E. coli</i>	B1	ATCC 25922	6.4	>25.6	>25.6	>25.6	>25.6	25.6	>25.6	3.2	$\leq$ 0.0125	1.6

<sup>a</sup> University of Leeds, Leeds, UK (I. Chopra).

<sup>b</sup> SENTRY study, University of Minden, Minden, Germany (F.J. Schmitz).

<sup>c</sup> Novartis AG, Vienna, Austria.

<sup>d</sup> University College London Hospitals, London, UK.

<sup>e</sup> General Hospital (AKH), Vienna, Austria.

Abbreviations: MSSA, methicillin sensitive *S. aureus*; MRSA, methicillin resistant *S. aureus*; MRSE, methicillin resistant *S. epidermidis*; VRSE, vancomycin resistant *S. epidermidis*; Clin. Isolate, clinical isolate; AZA, azamulin; AZM, azithromycin; DOX, doxycycline; LZD, linezolid; MXF, moxifloxacin; PLE, pleuromutilin; TIA, tiamulin; VAL, valnemulin.

TABLE 3. Antibacterial activity of BC-3311, BC-3291, conventional pleuromutilin derivatives and other standard antibiotics against *S. aureus* (MRSA)

<i>Antibiotic</i>	<i>n</i> <sup>a</sup>	<i>MIC</i> <sub>50</sub> ( $\mu\text{g/ml}$ )	<i>MIC</i> <sub>90</sub> ( $\mu\text{g/ml}$ )	<i>Range</i>	<i>% resistant</i> <sup>b</sup>
BC-3311	75	0.1	0.1	0.05-0.2	-
BC-3291	33	3.2	3.2	3.2-6.4	-
Pleuromutilin	33	1.6	1.6	0.8-1.6	-
Tiamulin	75	0.4	0.8	0.2-0.8	-
Azamulin	75	0.4	0.8	0.2-0.8	-
Valnemulin	75	0.05	0.1	$\leq 0.0125$ -0.1	-
Linezolid	75	2	2	1-4	0
Azithromycin	74	>256	>256	$\leq 0.0125$ ->256	76
Moxifloxacin	75	2	8	$\leq 0.125$ -16	59
Doxycycline	74	0.5	16	$\leq 0.0125$ -32	28

<sup>a</sup> number of isolates.

<sup>b</sup> percentage of strains resistant to the indicated antibiotic according to CLSI breakpoints.

TABLE 4. Summarized footprints of pleuromutilin derivatives in the central loop region of 23S rRNA domain V of *S. aureus* and *E. coli* ribosomes

<i>Nucleotide</i> ( <i>E. coli</i> <i>numbering</i> )	<i>Nucleotide</i> ( <i>S. aureus</i> <i>numbering</i> )	<i>Primer</i> <i>for</i> <i>S. aureus</i> / <i>E. coli</i>	<i>S. aureus</i> <sup>a</sup>	<i>E. coli</i> <sup>a</sup>	<i>Reference for</i> <i>E. coli</i>
A2058	A2085	P092-2Sa / P092	E	E	(25, 46)
A2059	A2086	P092-2Sa / P092	E	E	(25, 46)
U2192 <sup>b</sup>	C2219 <sup>b</sup>	P093Sa / P093	E	E	This study
U2506	U2533	P094Sa / P094	no P	P	(25, 31)
U2533	U2560	P095Sa / P095	P	no P	This study
U2563	U2590	P095Sa / P095	P	no P	This study
U2584	U2611	P095Sa / P095	P	P	(25, 31)
U2585	U2612	P095Sa / P095	P	P	(25, 31)

<sup>a</sup> E, enhancement; P, protection.

<sup>b</sup> Nucleotide at *E. coli* position 2192 is U in *E. coli* and C in *S. aureus*.

## 10. References

---

- Altman, S., Wesolowski, D., Guerrier-Takada, C., and Li, Y. (2005) RNase P cleaves transient structures in some riboswitches. *Proc Natl Acad Sci U S A* **102**: 11284-11289.
- Altuvia, S., and Wagner, E.G. (2000) Switching on and off with RNA. *Proc Natl Acad Sci U S A* **97**: 9824-9826.
- Altuvia, S. (2007) Identification of bacterial small non-coding RNAs: experimental approaches. *Curr Opin Microbiol* **10**: 257-261.
- Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G., Margalit, H., and Altuvia, S. (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* **11**: 941-950.
- Babitzke, P., and Romeo, T. (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol* **10**: 156-163.
- Bachler, M., Schroeder, R., and von Ahsen, U. (1999) StreptoTag: a novel method for the isolation of RNA-binding proteins. *Rna* **5**: 1509-1516.
- Baron, S., and National Center for Biotechnology Information (U.S.) (1996) Medical microbiology. Galveston, Tex.: University of Texas Medical Branch at Galveston.
- Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J.K., and Breaker, R.R. (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A* **101**: 6421-6426.
- Batey, R.T., Rambo, R.P., Lucast, L., Rha, B., and Doudna, J.A. (2000) Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* **287**: 1232-1239.
- Beckert, S., Kreikemeyer, B., and Podbielski, A. (2001) Group A streptococcal *rofA* gene is involved in the control of several virulence genes and eukaryotic cell attachment and internalization. *Infect Immun* **69**: 534-537.
- Bernstein, H.D., Poritz, M.A., Strub, K., Hoben, P.J., Brenner, S., and Walter, P. (1989) Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. *Nature* **340**: 482-486.
- Bisno, A.L. (1995) *Streptococcus pyogenes*. In Mandell G.L., Bennet, R.G., and Dolin, R. (eds.) *Principles and Practice of Infectious Diseases*, Vol. 2. Churchill Livingstone, New York, NY, pp. 1786-1799.
- Bisno, A.L., Brito, M.O., and Collins, C.M. (2003) Molecular basis of group A streptococcal virulence. *Lancet Infect Dis* **3**: 191-200.
- Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., Chevalier, C., Helfer, A.C., Benito, Y., Jacquier, A., Gaspin, C., Vandenesch, F., and Romby, P. (2007) *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev* **21**: 1353-1366.

- Brennan, R.G., and Link, T.M. (2007) Hfq structure, function and ligand binding. *Curr Opin Microbiol* **10**: 125-133.
- Brown, J., Farnsworth, R., Wannamaker, L.W., and Johnson, D.W. (1974) CAMP factor of group B streptococci: production, assay, and neutralization by sera from immunized rabbits and experimentally infected cows. *Infect Immun* **9**: 377-383.
- Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Smoot, L.M., Graham, M.R., Watson, R.O., and Musser, J.M. (2002) Rgg influences the expression of multiple regulatory loci to coregulate virulence factor expression in *Streptococcus pyogenes*. *Infect Immun* **70**: 762-770.
- Chen, Q., and Crosa, J.H. (1996) Antisense RNA, fur, iron, and the regulation of iron transport genes in *Vibrio anguillarum*. *J Biol Chem* **271**: 18885-18891.
- Choonee, N., Even, S., Zig, L., and Putzer, H. (2007) Ribosomal protein L20 controls expression of the *Bacillus subtilis* infC operon via a transcription attenuation mechanism. *Nucleic Acids Res* **35**: 1578-1588.
- Churchward, G. (2007) The two faces of Janus: virulence gene regulation by CovR/S in group A streptococci. *Mol Microbiol* **64**: 34-41.
- Christie, R., Atkins, N.E., Munch-Petersen, E.A. (1944) A note on a lytic phenomenon shown by group B streptococci. *Aust J Exp Biol Med Sci* **22**: 197-200.
- Cunningham, M.W. (2000) Pathogenesis of group A streptococcal infections. *Clin Microbiol Rev* **13**: 470-511.
- Cunningham, M.W. (2008) Pathogenesis of group A streptococcal infections and their sequelae. *Adv Exp Med Biol* **609**: 29-42.
- Deutscher, M.P. (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res* **34**: 659-666.
- El-Huneidi, W., Mui, R., Zhang, T.H., and Palmer, M. (2007) *Streptococcus agalactiae* CAMP factor/protein B does not bind to human IgG. *Med Microbiol Immunol* **196**: 73-77.
- Facklam, R.F., Martin, D.R., Lovgren, M., Johnson, D.R., Efstratiou, A., Thompson, T.A., Gowan, S., Kriz, P., Tyrrell, G.J., Kaplan, E., and Beall, B. (2002) Extension of the Lancefield classification for group A streptococci by addition of 22 new M protein gene sequence types from clinical isolates: emm103 to emm124. *Clin Infect Dis* **34**: 28-38.
- Ferretti, J.J., McShan, W.M., Ajdic, D., Savic, D.J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A.N., Kenton, S., Lai, H.S., Lin, S.P., Qian, Y., Jia, H.G., Najar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A., and McLaughlin, R. (2001) Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc Natl Acad Sci U S A* **98**: 4658-4663.
- Fischetti, V.A. (1991) Streptococcal M protein. *Sci Am* **264**: 58-65.
- Fogg, G.C., and Caparon, M.G. (1997) Constitutive expression of fibronectin binding in *Streptococcus pyogenes* as a result of anaerobic activation of rofA. *J Bacteriol* **179**: 6172-6180.



- Fogg, G.C., Gibson, C.M., and Caparon, M.G. (1994) The identification of *rofA*, a positive-acting regulatory component of *prtF* expression: use of an  $\gamma$  delta-based shuttle mutagenesis strategy in *Streptococcus pyogenes*. *Mol Microbiol* **11**: 671-684.
- Gase, K., Ferretti, J.J., Primeaux, C., and McShan, W.M. (1999) Identification, cloning, and expression of the CAMP factor gene (*cfa*) of group A streptococci. *Infect Immun* **67**: 4725-4731.
- Geisinger, E., Adhikari, R.P., Jin, R., Ross, H.F., and Novick, R.P. (2006) Inhibition of rot translation by RNAIII, a key feature of *agr* function. *Mol Microbiol* **61**: 1038-1048.
- Gelfand, M.S., Mironov, A.A., Jomantas, J., Kozlov, Y.I., and Perumov, D.A. (1999) A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet* **15**: 439-442.
- Gerdes, K., and Wagner, E.G. (2007) RNA antitoxins. *Curr Opin Microbiol* **10**: 117-124.
- Gonzalez, N., Heeb, S., Valverde, C., Kay, E., Reimann, C., Junier, T., and Haas, D. (2008) Genome-wide search reveals a novel GacA-regulated small RNA in *Pseudomonas* species. *BMC Genomics* **9**: 167.
- Gottesman, S. (2004) The small RNA regulators of *Escherichia coli*: roles and mechanisms\*. *Annu Rev Microbiol* **58**: 303-328.
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**: D121-124.
- Grundy, F.J., and Henkin, T.M. (2003) The T box and S box transcription termination control systems. *Front Biosci* **8**: d20-31.
- Gryllos, I., Levin, J.C., and Wessels, M.R. (2003) The CsrR/CsrS two-component system of group A *Streptococcus* responds to environmental Mg<sup>2+</sup>. *Proc Natl Acad Sci U S A* **100**: 4227-4232.
- Guillier, M., Allemand, F., Raibaud, S., Dardel, F., Springer, M., and Chiaruttini, C. (2002) Translational feedback regulation of the gene for L35 in *Escherichia coli* requires binding of ribosomal protein L20 to two sites in its leader mRNA: a possible case of ribosomal RNA-messenger RNA molecular mimicry. *Rna* **8**: 878-889.
- Herskovits, A.A., Bochkareva, E.S., and Bibi, E. (2000) New prospects in studying the bacterial signal recognition particle pathway. *Mol Microbiol* **38**: 927-939.
- Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, S., Tacker, M. and Schuster, P. (1994) Fast Folding and Comparison of RNA Secondary Structures. *Monatshefte f. Chemie* **125**, 167-188.
- Hondorp, E.R., and McIver, K.S. (2007) The Mga virulence regulon: infection where the grass is greener. *Mol Microbiol* **66**: 1056-1065.
- Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C., Jacquier, A., Vandenesch, F., and Romby, P. (2005) *Staphylococcus aureus* RNAIII and

- the endoribonuclease III coordinately regulate spa gene expression. *Embo J* **24**: 824-835.
- Huttenhofer, A., and Vogel, J. (2006) Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res* **34**: 635-646.
- Hynes, W. (2004) Virulence factors of the group A streptococci and genes that regulate their expression. *Front Biosci* **9**: 3399-3433.
- Johansson, J., and Cossart, P. (2003) RNA-mediated control of virulence gene expression in bacterial pathogens. *Trends Microbiol* **11**: 280-285.
- Kawano, M., Reynolds, A.A., Miranda-Rios, J., and Storz, G. (2005) Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res* **33**: 1040-1050.
- Kazantsev, A.V., and Pace, N.R. (2006) Bacterial RNase P: a new view of an ancient enzyme. *Nat Rev Microbiol* **4**: 729-740.
- Keiler, K.C. (2007) Physiology of tmRNA: what gets tagged and why? *Curr Opin Microbiol* **10**: 169-175.
- Klenk, M., Koczan, D., Guthke, R., Nakata, M., Thiesen, H.J., Podbielski, A., and Kreikemeyer, B. (2005) Global epithelial cell transcriptional responses reveal *Streptococcus pyogenes* Fas regulator activity association with bacterial aggressiveness. *Cell Microbiol* **7**: 1237-1250.
- Kreikemeyer, B., Boyle, M.D., Buttaro, B.A., Heinemann, M., and Podbielski, A. (2001) Group A streptococcal growth phase-associated virulence factor regulation by a novel operon (Fas) with homologies to two-component-type regulators requires a small RNA molecule. *Mol Microbiol* **39**: 392-406.
- Kreikemeyer, B., McIver, K.S., and Podbielski, A. (2003) Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. *Trends Microbiol* **11**: 224-232.
- Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S., and Bassler, B.L. (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* **118**: 69-82.
- Liu, J.M., Bittker, J.A., Lonshteyn, M., and Liu, D.R. (2005) Functional dissection of sRNA translational regulators by nonhomologous random recombination and in vivo selection. *Chem Biol* **12**: 757-767.
- Livny, J., Brencic, A., Lory, S., and Waldor, M.K. (2006) Identification of 17 *Pseudomonas aeruginosa* sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. *Nucleic Acids Res* **34**: 3484-3493.
- Majdalani, N., Hernandez, D., and Gottesman, S. (2002) Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol Microbiol* **46**: 813-826.
- Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C., and Breaker, R.R. (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**: 577-586.
- Mandal, M., and Breaker, R.R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**: 29-35.

- Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L., and Breaker, R.R. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**: 275-279.
- Mandin, P., Repoila, F., Vergassola, M., Geissmann, T., and Cossart, P. (2007) Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res* **35**: 962-974.
- Mangold, M., Siller, M., Roppenser, B., Vlamincx, B.J., Penfound, T.A., Klein, R., Novak, R., Novick, R.P., and Charpentier, E. (2004) Synthesis of group A streptococcal virulence factors is controlled by a regulatory RNA molecule. *Mol Microbiol* **53**: 1515-1527.
- Masse, E., and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**: 4620-4625.
- Masse, E., Escorcia, F.E., and Gottesman, S. (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev* **17**: 2374-2383.
- McIver, K.S., and Scott, J.R. (1997) Role of *mga* in growth phase regulation of virulence genes of the group A streptococcus. *J Bacteriol* **179**: 5178-5187.
- Merianos, H.J., Wang, J., and Moore, P.B. (2004) The structure of a ribosomal protein S8/spc operon mRNA complex. *Rna* **10**: 954-964.
- Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A., and Nudler, E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**: 747-756.
- Moller, T., Franch, T., Udesen, C., Gerdes, K., and Valentin-Hansen, P. (2002) Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev* **16**: 1696-1706.
- Nahvi, A., Barrick, J.E., and Breaker, R.R. (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res* **32**: 143-150.
- Ontiveros-Palacios, N., Smith, A.M., Grundy, F.J., Soberon, M., Henkin, T.M., and Miranda-Rios, J. (2008) Molecular basis of gene regulation by the THI-box riboswitch. *Mol Microbiol* **67**: 793-803.
- Peck-Miller, K.A., and Altman, S. (1991) Kinetics of the processing of the precursor to 4.5 S RNA, a naturally occurring substrate for RNase P from *Escherichia coli*. *J Mol Biol* **221**: 1-5.
- Phan, T.T., and Schumann, W. (2007) Development of a glycine-inducible expression system for *Bacillus subtilis*. *J Biotechnol* **128**: 486-499.
- Pritchard, K.H., and Cleary, P.P. (1996) Differential expression of genes in the vir regulon of *Streptococcus pyogenes* is controlled by transcription termination. *Mol Gen Genet* **250**: 207-213.
- Ribardo, D.A., and McIver, K.S. (2006) Defining the Mga regulon: Comparative transcriptome analysis reveals both direct and indirect regulation by Mga in the group A streptococcus. *Mol Microbiol* **62**: 491-508.
- Rivas, E., and Eddy, S.R. (2001) Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2**: 8.

- Rivas, E., Klein, R.J., Jones, T.A., and Eddy, S.R. (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* **11**: 1369-1373.
- Roberts, S.A., and Scott, J.R. (2007) RivR and the small RNA RivX: the missing links between the CovR regulatory cascade and the Mga regulon. *Mol Microbiol* **66**: 1506-1522.
- Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. (2002) Comparative genomics of thiamin biosynthesis in prokaryotes. New genes and regulatory mechanisms. *J Biol Chem* **277**: 48949-48959.
- Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. (2003) Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Res* **31**: 6748-6757.
- Romby, P., Vandenesch, F., and Wagner, E.G. (2006) The role of RNAs in the regulation of virulence-gene expression. *Curr Opin Microbiol* **9**: 229-236.
- Romeo, T., Gong, M., Liu, M.Y., and Brun-Zinkernagel, A.M. (1993) Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J Bacteriol* **175**: 4744-4755.
- Skalka, B., and Smola, J. (1981) Lethal effect of CAMP-factor and UBERIS-factor--a new finding about diffusible exo-substances of streptococcus agalactiae and Streptococcus uberis. *Zentralbl Bakteriol A* **249**: 190-194.
- Sorek, R., Kunin, V., and Hugenholtz, P. (2008) CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**: 181-186.
- Sriskandan, S., Faulkner, L., and Hopkins, P. (2007) Streptococcus pyogenes: Insight into the function of the streptococcal superantigens. *Int J Biochem Cell Biol* **39**: 12-19.
- Steiner, K., and Malke, H. (2001) *relA*-Independent amino acid starvation response network of Streptococcus pyogenes. *J Bacteriol* **183**: 7354-7364.
- Storz, G., Opdyke, J.A., and Zhang, A. (2004) Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol* **7**: 140-144.
- Storz, G., Altuvia, S., and Wassarman, K.M. (2005) An abundance of RNA regulators. *Annu Rev Biochem* **74**: 199-217.
- Storz, G., and Haas, D. (2007) A guide to small RNAs in microorganisms. *Curr Opin Microbiol* **10**: 1-3
- Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S., and Breaker, R.R. (2003) An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev* **17**: 2688-2697.
- Svenningsen, S.L., Waters, C.M., and Bassler, B.L. (2008) A negative feedback loop involving small RNAs accelerates *Vibrio cholerae*'s transition out of quorum-sensing mode. *Genes Dev* **22**: 226-238.
- Toledo-Arana, A., Repoila, F., and Cossart, P. (2007) Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol* **10**: 182-188.

- Tu, K.C., and Bassler, B.L. (2007) Multiple small RNAs act additively to integrate sensory information and control quorum sensing in *Vibrio harveyi*. *Genes Dev* **21**: 221-233.
- Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., and Gelfand, M.S. (2002) Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res* **30**: 3141-3151.
- Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., and Gelfand, M.S. (2003) Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *Rna* **9**: 1084-1097.
- Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., and Gelfand, M.S. (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet* **20**: 44-50.
- Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jager, J.G., Huttenhofer, A., and Wagner, E.G. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* **31**: 6435-6443.
- Vogl, C., Grill, S., Schilling, O., Stulke, J., Mack, M., and Stolz, J. (2007) Characterization of riboflavin (vitamin B2) transport proteins from *Bacillus subtilis* and *Corynebacterium glutamicum*. *J Bacteriol* **189**: 7367-7375.
- Voyich, J.M., Sturdevant, D.E., Braughton, K.R., Kobayashi, S.D., Lei, B., Virtaneva, K., Dorward, D.W., Musser, J.M., and DeLeo, F.R. (2003) Genome-wide protective response used by group A *Streptococcus* to evade destruction by human polymorphonuclear leukocytes. *Proc Natl Acad Sci U S A* **100**: 1996-2001.
- Wakeman, C.A., Winkler, W.C., and Dann, C.E., 3rd (2007) Structural features of metabolite-sensing riboswitches. *Trends Biochem Sci* **32**: 415-424.
- Washietl, S., and Hofacker, I.L. (2004) Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. *J Mol Biol* **342**: 19-30.
- Washietl, S., Hofacker, I.L., and Stadler, P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A* **102**: 2454-2459.
- Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* **15**: 1637-1651.
- Wassarman, K.M., and Saecker, R.M. (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* **314**: 1601-1603.
- Wassarman, K.M. (2007) 6S RNA: a regulator of transcription. *Mol Microbiol* **65**: 1425-1431.
- Weilbacher, T., Suzuki, K., Dubey, A.K., Wang, X., Gudapaty, S., Morozov, I., Baker, C.S., Georgellis, D., Babitzke, P., and Romeo, T. (2003) A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol Microbiol* **48**: 657-670.

- Wilderman, P.J., Sowa, N.A., FitzGerald, D.J., FitzGerald, P.C., Gottesman, S., Ochsner, U.A., and Vasil, M.L. (2004) Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc Natl Acad Sci U S A* **101**: 9792-9797.
- Windbichler, N., and Schroeder, R. (2006) Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nat Protoc* **1**: 637-640.
- Winkler, W., Nahvi, A., and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**: 952-956.
- Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E., and Breaker, R.R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* **10**: 701-707.
- Winkler, W.C., and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* **59**: 487-517.
- Withey, J.H., and Friedman, D.I. (2002) The biological roles of trans-translation. *Curr Opin Microbiol* **5**: 154-159.
- Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-3415.

Rfam database: Genome distribution: *Streptococcus pyogenes* M1 GAS, complete genome. [http://www.sanger.ac.uk/cgi-in/Rfam/genome\\_view.pl?genome\\_acc=AE004092](http://www.sanger.ac.uk/cgi-in/Rfam/genome_view.pl?genome_acc=AE004092), Ref. 23.05.2008

NCBI database: <http://www.ncbi.nlm.nih.gov/sites/entrez/>, Ref. 23.05.2008

## 11. Acknowledgements

---

I am very grateful to Dr. Emmanuelle Charpentier for giving me the opportunity to work in her laboratory. I would like to thank her as well for her outstanding supervision, support and encouragement during the last year.

I would like to thank Maria Eckert for supervising and supporting me very kindly the whole time. I would like to thank Karine Gonzales for critically reading the thesis manuscript and for her help with the figures that had to be changed.

My very special thank goes to my lab colleagues and friends Fanny Beneyt, Maria Eckert, Stefanie Füreder, Karine Gonzales and Silvia Spieß for the excellent advice concerning lab work and for providing me comfort and a warm and bright atmosphere during my time in the lab.

Most appreciation and respect goes to my family, my father, my mother, my sister Tina, and my grandmother, for their love and their support every single day and for always believing in me.

## 12. Curriculum Vitae

---

### Personal details

---

Name: Doris Veit

Date of birth: 10.09.1983

Place of birth: Wels

Nationality: Austria



### Education

---

2002-2008

University of Vienna

Molecular Biology

Main subjects: Microbiology, Cell biology and Developmental Biology

Diploma thesis: "Novel small non-coding RNAs in *Streptococcus pyogenes*"

(Max F. Perutz Laboratories, Department of Microbiology and

Immunology, Dr. Emmanuelle Charpentier's Laboratory)

1994-2002

Bundesgymnasium und Bundesrealgymnasium Dr. Schauerstraße, 4600

Wels

### Personal skills and competences

---

Mother tongue: German

Foreign languages: English (fluent in spoken and written)

French (basic knowledge)

Computer skills

Driving licence



# 13. Lebenslauf

---

## Persönliche Daten

---

Name: Doris Veit

Geburtsdatum: 10.09.1983

Geburtsort: Wels

Nationalität: Österreich



## Schul- und Berufsausbildung

---

2002-2008

Universität Wien

Studienrichtung: Molekularbiologie

Schwerpunktfächer: Mikrobiologie, Zellbiologie and Entwicklungsbiologie

Diplomarbeit: "Novel small non-coding RNAs in *Streptococcus pyogenes*"

(Max F. Perutz Laboratories, Department of Microbiology and Immunology, Dr. Emmanuelle Charpentier's Laboratory)

1994-2002

Bundesgymnasium und Bundesrealgymnasium Dr. Schauerstraße, 4600  
Wels

## Persönliche Fähigkeiten und Kompetenzen

---

Muttersprache: Deutsch

Fremdsprachen: Englisch (flüssig in Wort und Schrift)

Französisch (Grundkenntnisse)

EDV Kenntnisse

Führerschein B