Predicting small RNA targets in prokaryotes – a challenge beyond the barriers of thermodynamic models

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Technischen Fakultät
der Albert-Ludwigs-Universität Freiburg

2016

von

Diplom Biologe
Patrick R. Wright

Dekan:

Prof. Dr. Oliver Paul

Gutachter:

Prof. Dr. Rolf Backofen

Prof. Dr. Wolfgang R. Hess

Kommissionsvorsitz:

Prof. Dr. Christoph Scholl

Kommissionsbeisitz:

Prof. Dr. Gerald Urban

Datum der Promotion:

19.12.2016

“Ich möchte mein Leben so leben, dass die Welt dadurch jeden Tag ein wenig besser wird.”

– Melanie Stockhausen

Abstract

Advanced high throughput sequencing of nucleic acid samples coupled with sophisticated analysis algorithms has led to the discovery of a plethora of commonly non-coding but functional short transcripts present in diverse prokaryotic genomes. These so called *trans*-acting small RNAs fulfill their regulatory role by directly pairing with their targets via RNA-RNA interaction. Thus, this type of regulation is termed post-transcriptional. The small RNAs can impose both up- and down regulation of targets. While the discovery of thus far unknown transcripts has become a standardized procedure, the recovery of their regulatory targets has remained challenging. For this reason RNA-RNA interaction prediction algorithms that attempt target identification at the genomic scale have been developed. Leading methods include a thermodynamic energy model and incorporate interaction site accessibility for RNA duplex prediction. These methods are generally successful at predicting the correct RNA duplex for two RNA molecules known to interact, but fail at correctly predicting targets at the whole genome level due to high false positive prediction rates. For this reason the methods presented in this thesis attempt to employ and extend the purely thermodynamic predictors. Indeed, the inclusion of phylogenetic information, pathway analysis and whole genome binding data for the RNA chaperone Hfq allow a significant reduction of false positives in the in silico prediction lists. Systematic benchmarking revealed that the newly developed algorithm CopraRNA not only outperforms other small RNA target predictors, but also rivals microarray driven experimental target prediction. Next to the convincing benchmark performance on already known RNA-RNA interactions, 23 novel, previously unreported and promising target candidates from *Escherichia coli* were retrieved from the CopraRNA predictions lists. Of these, 17 were experimentally confirmed in the wet-lab. This shows that the algorithm is also able to produce qualitatively new insights. Follow up studies on the rhizobial AbcR1 and EcpR1 small RNAs furthermore validate the applicability of CopraRNA beyond enterobacterial species. The algorithm has been made available as an easy to use web server interface and is being amply accessed by the research community. Finally, recent and currently unpublished results indicate that an extended version of CopraRNA may be able to detect hot spots of evolutionary diversity in post-transcriptional gene regulatory networks.

Keywords

bacteria – target prediction – sRNA – RNA-RNA interaction – RNA binding protein – post-transcriptional gene regulation

Zusammenfassung

Fortschrittliche Hochdurchsatzmethoden zur Sequenzierung von Nukleinsäureseproben gekoppelt mit anspruchsvollen Analysealgorithmen haben zur Entdeckung einer Vielzahl kleiner, funktionaler und nichtkodierender Transkripte in prokaryotischen Genomen geführt. Diese sogenannten *trans*-agierenden kleinen RNAs vollziehen ihre regulatorischen Rollen indem sie über RNA-RNA Interaktion direkt an ihre Ziel-RNAs binden, weshalb dieser Regulationsmechanismus als post-transkriptionell bezeichnet wird. Die kleinen RNAs können auf ihre regulatorischen Ziele sowohl einen aktivierenden als auch deaktivierenden Effekt ausüben. Während die Entdeckung von neuen Transkripten mittlerweile standardisiert ist, gehört die Suche nach Ziel-RNAs nach wie vor zu den anspruchsvollen Aufgaben in diesem Forschungsfeld. Aus diesem Grund sind in der Vergangenheit Algorithmen entwickelt worden, die genomweite Zielvorhersagen ermöglichen. Führende Ansätze verwenden dafür thermodynamische Energiemodelle und beinhalten des Weiteren eine Betrachtung der Zugänglichkeit der RNA Interaktionsstellen. Für RNA Moleküle von denen bekannt ist, dass sie interagieren, können diese Algorithmen meist die korrekte Interaktion vorhersagen. Problematischer sind genomweite Vorhersagen, da der Anteil falsch positiver Vorhersagen nachwievor zu groß ist. Aus diesem Grund versuchen die Methoden, die in dieser Arbeit vorgestellt werden, die eben genannten thermodynamischen Ansätze anzuwenden und zu erweitern. In der Tat hat sich gezeigt, dass die Verwendung von phylogenetischer Information, funktioneller Anreicherungsanalysen und genomweiter Bindekarten des RNA Chaperons Hfq zu einer signifikanten Reduktion von falsch positiven Vorhersagen führen kann. Systematische Benchmarkanalysen haben gezeigt, dass der neu entwickelte CopraRNA Algorithmus nicht nur bisher führende Vorhersagealgorithmen übertrifft, sondern auch mit experimentellen microarraybasierten Zielvorhersagen konkurrieren kann. Neben den überzeugenden Benchmarkergebnissen für bekannte RNA-RNA Interaktionen, wurden 23 weitere, von CopraRNA vorhergesagte, bisher unbekannte RNA-RNA Interaktionen aus *Escherichia coli* im Labor untersucht. Von diesen konnten 17 verifiziert werden. Dies zeigt, dass der Algorithmus auch qualitativ neue Erkenntnisse hervorbringen kann. Nachfolgestudien mit den rhizobiellen RNAs AbcR1 und EcpR1 haben des Weiteren die Anwendbarkeit von CopraRNA über Enterobakterien hinaus bestätigt. Für den Algorithmus ist ein Webserver Interface entwickelt worden, welches rege Anwendung findet. Abschließend ist darauf hinzuweisen, dass vorläufige, unpublizierte Ergebnisse aus der aktuellen Entwicklung und Forschung darauf hinweisen, dass eine erweiterte Version von CopraRNA dazu im Stande sein könnte “hot spots” evolutionärer Diversität in post-transkriptionellen regulatorischen Netzwerken vorherzusagen.

Personal publication list

Nicholas J. Tobias, Antje K. Heinrich, Helena Eresmann, **Patrick R. Wright**, Nick Neubacher, Rolf Backofen and Helge B. Bode (2016) *Photorhabdus*-nematode symbiosis is dependent on *hfq*-mediated regulation of secondary metabolites. **Environmental microbiology**

Erik Holmqvist, **Patrick R. Wright**, Lei Li, Thorsten Bischler, Lars Barquist, Richard Reinhardt, Rolf Backofen and Jörg Vogel (2016) Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*. **The EMBO Journal**, 35, 991-1011

Kehau Hagiwara[†], **Patrick R. Wright**[†], Nicole K. Tabandera, Dovi Kelman, Rolf Backofen, Sesselja Ómarsdóttir and Anthony D. Wright (2015) Comparative analysis of the antioxidant properties of Icelandic and Hawaiian lichens. **Environmental microbiology**, 18, 2319-2325, [†] Shared first authors

Marta Robledo, Benjamin Frage, **Patrick R. Wright** and Anke Becker (2015) A stress-induced small RNA modulates alpha-rhizobial cell cycle progression. **PLOS Genetics**, 11, e1005153

Patrick R. Wright[†], Jens Georg[†], Martin Mann[†], Dragos A. Sorescu, Andreas S. Richter, Steffen Lott, Robert Kleinkauf, Wolfgang R. Hess and Rolf Backofen (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. **Nucleic acids research**, 42, W119-W123 [†] Shared first authors

Tomasz Chelmicki, Friederike Dündar, Matthew James Turley, Tasneem Khanam, Tugce Aktas, Fidel Ramírez, Anne-Valerie Gendrel, **Patrick R. Wright**, Pavankumar Videm, Rolf Backofen, Edith Heard, Thomas Manke and Asifa Akhtar (2014) MOF-associated complexes ensure stem cell identity and Xist repression. **eLife**, 3, e02024

Aaron Overlöper, Alexander Kraus, Rosemarie Gurski, **Patrick R. Wright**, Jens Georg, Wolfgang R. Hess and Franz Narberhaus (2014) Two separate modules of the conserved regulatory RNA AbcR1 address multiple target mRNAs in and outside of the translation initiation region. **RNA Biology**, 11, 624-640

Patrick R. Wright, Andreas S. Richter, Kai Papenfort, Martin Mann, Jörg Vogel, Wolfgang R. Hess, Rolf Backofen and Jens Georg (2013) Comparative genomics boosts target prediction for bacterial small RNAs. **Proceedings of the National Academy of Sciences**, 110, E3487-E3496

Stephan Felder, Sandra Dreisigacker, Stefan Kehraus, Edith Neu, Gabriele Bierbaum, **Patrick R. Wright**, Dirk Menche, Till F. Schäberle and Gabriele M. König (2013) Salimabromide: Unexpected chemistry from the obligate marine myxobacterium *Enhygromxya salina*. **Chemistry-A European Journal**, 19, 9319-9324

Dovi Kelman, Ellen Kromkowski Posner, Karla J. McDermid, Nicole K. Tabandera, **Patrick R. Wright** and Anthony D. Wright (2012) Antioxidant activity of Hawaiian marine algae. **Marine drugs**, 10, 403-416

Rahul V. Haware, **Patrick R. Wright**, Kenneth R. Morris and Mazen L. Hamad (2011) Data fusion of Fourier transform infrared spectra and powder X-ray diffraction patterns for pharmaceutical mixtures. **Journal of pharmaceutical and biomedical analysis**, 56, 944-949

Oral presentations

Patrick R. Wright (2016) Recent progress in the in silico analysis of small RNA networks. Sensory and regulatory RNAs in Prokaryotes, München, Germany

Patrick R. Wright (2015) Towards exhaustive target prediction for prokaryotic *trans* acting small RNAs. Sensory and regulatory RNAs in Prokaryotes and CRISPR-Cas, Braunschweig, Germany

Patrick R. Wright (2013) Comparative genomics boosts target prediction for bacterial small RNAs. 28th TBI winter seminar, Bled, Slovenia

Poster presentations

Patrick R. Wright, Martin Mann, Robert Kleinkauf, Dragos A. Sorescu, Sita J. Lange, Torsten Houwaart, Omer S. Alkhnbashi, Dominic Rose, Steffen Heyne, Andreas S. Richter, Wolfgang R. Hess, Jens Georg, Anke Busch, Sebastian Will and Rolf Backofen (2015) Freiburg RNA tools webserver. Regulating with RNA in Bacteria and Archaea, Cancun, Mexico

Patrick R. Wright, Andreas S. Richter, Kai Papenfort, Jörg Vogel, Wolfgang R. Hess, Jens Georg and Rolf Backofen (2012) Homology IntaRNA (hintaRNA) - Comparative prediction of sRNA targets in prokaryotes. Sensory and regulatory RNAs in Prokaryotes, Bochum, Germany

Aaron Overlöper, Philip Möller, Ina Wilms, Jens Georg, **Patrick R. Wright**, Wolfgang R. Hess and Franz Narberhaus (2012) AbcR1 sRNA regulates multiple targets in *Agrobacterium tumefaciens*. Sensory and regulatory RNAs in Prokaryotes, Bochum, Germany

Contents

1 Introduction

In traditional science, researchers would generally perform single, small scale experiments to investigate their hypotheses. Modern technological and scientific advances have enabled approaches in which millions of experiments can be carried out simultaneously. These experimental techniques have elevated life sciences to a new level by yielding ever increasing volumes of data that call for automated analysis algorithms. This has led to the establishment of the research field which is now referred to as computational biology or alternatively bioinformatics. Bioinformatics is an interdisciplinary field that analyses data from many biological subjects. Among others, it allows interpretation of genomic, transcriptomic, proteomic and metabolomic data. Often, data from different sources need to be connected to generate an integrated understanding of the analyzed biological system.

Transcriptomics, the term referring to high throughput sequencing of transcribed RNA molecules, can be employed to retrieve a snapshot of what cells' transcriptomic output looks like at a specified time point and condition. During the last decade, transcriptomics studies have unveiled a plethora of previously unknown RNA-based regulators of cellular physiology. Many of these regulators perform their functional tasks via direct RNA-RNA interactions. The importance of such RNA-RNA interaction driven regulation both in eukaryotes and prokaryotes has thus become clear. While the experimental discovery of new RNA-based regulators is now readily possible, the retrieval of their regulatory targets in the wet-lab remains laborious. For this reason algorithms, which attempt in silico target prediction at the genomic scale, have been developed. Advanced algorithms employ thermodynamic energy models to perform their predictions, but still suffer from high false positives rates. The work presented in this thesis focuses on the development of methods that enhance the thermodynamic models in order to significantly reduce the amount of false positive predictions in whole genome target predictions for prokaryotic *trans*-acting small RNAs. The results show that the newly introduced approaches represent a real alternative and extension to purely wet-lab-based predictions and are generically applicable also beyond model Enterobacteriaceae.

1.1 Structure of the thesis

This thesis is composed of several parts. Firstly, an introduction is provided, which outlines the biological background and the algorithmic setting. After the introductory part, the subsequent chapters consist of a discussion and an outlook on the future of the field. The final chapters contain the research papers published during the completion of this thesis.

1.2 DNA, RNA, proteins and the central dogma of molecular biology

In the early 1950s, Rosalind Franklin produced high quality X-ray diffraction photographs of <u>d</u>eoxyribo<u>n</u>ucleic <u>a</u>cid[1]. Subsequently, these images aided James Watson and Francis Crick in unveiling the double helical structure of this macromolecule, now commonly referred to as DNA[2]. DNA is a prevalently double stranded anti-parallel molecule consisting of a sugar phosphate backbone with organic bases extending towards the molecule's central axis. The backbones of the two strands are oriented in a 5' to 3' direction for one of the DNA strands and in a 3' to 5' direction for its partner. The 5' and 3' pinpoint positions on the backbone's ribose portion. In this notation, 5' refers to the 5' phosphate group and 3' refers to the 3' hydroxyl moiety (see Figure 1). The DNA's bases are adenine (A), guanine (G), cytosine (C) and thymine (T). The three components; phosphate, pentose and nucleobase make up a nucleotide (see Figure 1). The forces holding the two strands together are hydrogen bonds. These bonds form between either A and T pairs, which involve two hydrogen bonds, or between G and C pairs, which are connected by three hydrogen bonds and are thus more stable than the A-T pairs. Hence, the strands are complementary to each other, and the nucleotide sequence of one strand infers the nucleotide sequence of its anti-parallel counterpart. Taking this property into account, Watson and Crick also elegantly deduced DNA's inherent ability to serve as a template for its own replication[2] (see Figure 2). Over two decades after Watson and Crick proposed their model for DNA's chemical structure, Frederick Sanger published a method to precisely and efficiently determine the nucleotide sequence of DNA molecules[3]. This so called chain-termination method was employed until next generation sequencing (NGS) methods became common place in the middle of the 2000s[4]. One of the first NGS methods available is referred to as pyrosequencing. This method specifically exploits the release of pyrophosphate during the process of DNA polymerization[5,6].

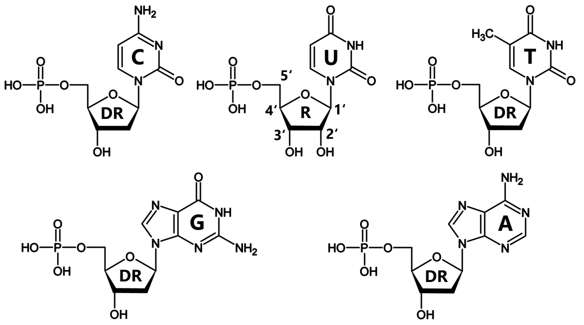


Figure 1. The image shows the chemical structure of the four possible DNA nucleotides and one RNA nucleotide. A nucleotide is composed of three central components. Firstly, the base portion, which can be a pyrimidine or a purine. The pyrimidines are cytosine (C), thymine (T) and uracil (U). The purines are guanine (G) and adenine (A). Secondly, the pentose which is deoxyribose (DR) in DNA and ribose (R) in RNA. The third component is the phosphate on the left side of the structures. The numbering of the carbon atoms on the ribose ring from 1' to 5' is indicated on the U containing nucleotide.

Currently, the most common NGS method is referred to as Illumina sequencing. In contrast to pyrosequencing, it does not utilize released pyrophosphate, but rather employs fluorescently tagged nucleotides to asses the sequence composition of a given DNA sample[7]. Both aforementioned NGS methods apply the general principle of "sequencing by synthesis"[4]. These breakthroughs in sequence identification marked the advent of modern biological research, which strongly focuses on the analysis of nucleic acid sequences.

Next to the widely known DNA, ribonucleic acid (RNA) is also a key player in biological networks and takes a prime position in current life science related research projects. In contrast to DNA, RNA is made up of the basic building blocks A, G, C and uracil (U) (see Figure 1) instead of T, and is biologically produced from a DNA template in a process known as transcription (see Figure 2). Furthermore, its sugar phosphate backbone differs slightly from that of DNA in that it includes an additional hydroxyl group at the 2'-position of the backbone's pentose ring (see Figure 1). This is also the reason for the naming of the two molecules. DNA is referred to as "deoxy" because it does not contain this hydroxyl group in its backbone. Intriguingly, and in contrast to the most common current flow of genetic information (DNA → RNA), it is assumed that RNA preceded DNA in what is referred to as the "RNA world"[8]. In this setting, RNA is believed to have performed catalytic reactions independently of proteins until RNA molecules evolved that permitted protein synthesis via translation of RNA. While proteins have taken over many catalytic tasks, some RNA types are still catalysts for biochemical reactions. These RNAs are referred to as ribozymes[9]. The increasing complexity eventually led to today's world in which the central dogma of molecular biology governs the transfer of genetic information between the macromolecular classes DNA, RNA and protein[10] (see Figure 2) under application of the mostly universal genetic code[11].

For a long time, RNA was mainly considered as a messenger molecule (mRNA) between DNA and proteins. Furthermore, only approximately one percent of the human genome encodes for proteins[12]. This misled scientists into proposing models in which the majority of genomic DNA had no function since large portions of this DNA do not code for proteins. Employing the previously mentioned NGS methods, the sequencing of RNA (RNA-seq) molecules at the genomic scale has become possible[13,14], and many functional transcripts not coding for proteins have since been identified and characterized, thus inferring functionality for most of the genomic DNA[15].

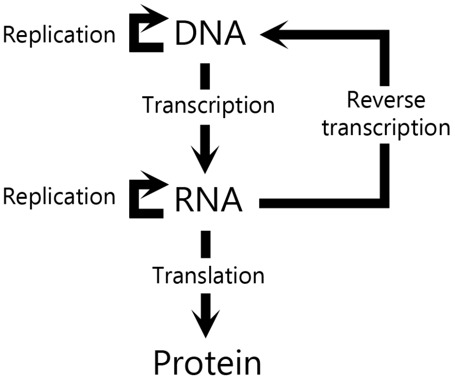


Figure 2. The scheme shows the directionality of genetic information as dictated by the central dogma of molecular biology. Reverse transcription and RNA replication are processes performed by some viruses[16,17]. Reverse transcription can also be performed by telomerase[18].

In this context, a complex RNA language has been suggested[19]. Lately, RNA-seq has been extended to single cell resolution, which allows more detailed investigations compared to the sequencing data retrieved from bulk samples[20,21].

In the investigation of prokaryotic organisms, RNAs encoded in genomic regions between protein coding genes have become the focus of considerable interest in recent times[22–25]. Analysis of bacterial RNA sequencing data and characterization of novel RNA-based regulators mark the core of the research described in this thesis.

1.3 Bacteria

Bacteria are a phylogenetically diverse subgroup of prokaryotic organisms. For practical reasons, they can be grouped into Gram-positive and Gram-negative species. Gram-positive bacteria have a thicker cell wall when compared to Gram-negative cells. The Gram classification reflects the different species' reaction to the Gram staining protocol[26]. In this test, Gram-negative specimens appear pink, while Gram-positive bacteria are purple. Next to the cell wall, bacteria commonly feature several additional cellular structures. These are, the capsule, the plasma membrane, the cytoplasm, ribosomes, flagella, pili, plasmids and the nucleoid. Both, plasmids and the chromosomal DNA in the nucleoid contain the bacterial cell's genetic information. They are usually circular but linear structures also exist[27]. Further cellular components such as gas vesicles[28] or storage grains have been reported[29].

Different bacterial species exhibit countless cellular morphologies and shapes[30] and often bacteria are observed as singular cells. Yet, several examples of multicellular bacterial associations are documented[31]. A popular example for multicellular bacteria are cyanobacteria of the genus *Anabaena*, which form filaments of connected cells that may even be differentiated into distinct, specialized cell types called heterocysts[32,33].

The cell cycle is the central model for bacterial reproduction and consists of several stages[34]. It starts with a newly divided cell which begins replicating its genetic material. After this replication phase, the replicons (i.e. chromosomes and plasmids) start segregating to the cell poles. Furthermore, FtsZ proteins start forming a ring structure at the center of the cell, the so called Z-ring[35]. The Z-ring serves as infrastructure for the recruitment of other cell division factors[36]. After the replicons have successfully segregated, the bacterial cell constricts in the middle and divides, thus restarting the cell cycle from the beginning. The study presented in Chapter 7 outlines how a bacterial RNA-based regulator (EcpR1) can influence cell cycle dynamics.

The cycle governs bacterial growth, which itself can also be subdivided into several phases[37]. The first phase is the lag phase in which bacteria adjust their metabolism to the environment. This is followed by the log phase in which bacterial cells proliferate exponentially. For this reason it is also referred to as the exponential phase. Upon onset of nutrient depletion, bacteria transition into the stationary phase, where the rate of cell death equals the rate of newly spawned cells and thus the population ceases to increase. The final phase is the death phase, where the majority of bacteria start dying due to a lack of nutrients. In order to overcome death due to such nutrient depletion, certain species have evolved to form endospores[38]. These endospores are resilient to lack of nutrients, heat, radiation and reactive chemicals[39]. Furthermore, it has been claimed that endospores may be able to survive for millions of years in this durable state[40].

Bacteria can be considered as a strikingly important, successful and adaptive group of organisms that has significant influence on all ecosystems and hence detailed investigation of these organisms is vital. Cyanobacteria for instance play a central role in global primary production and oxygen turnover[41] and are also the origin of chloroplasts which enable photosynthesis in higher plants[42]. A recent study was even able to identify a "cyanobacterial eye" which aids *Synechocystis* sp. PCC 6803 in phototaxis[43]. Moreover, bacterial communities have been shown to communicate using chemical signaling molecules. This process, known as quorum sensing[44,45], is important for bacterial communities to enable them to asses local population density and make collective metabolic decisions. If the population density is low, it makes no sense to produce exoenzymes or induce fluorescence. On the other hand, dense populations of bacterial cells favor the initiation of such cooperative processes.

The adaptive nature of bacteria is best explained by considering the clinical setting where they cause human disease and in severe cases death. After the discovery[46] and successful application of penicillin, it was originally believed that bacterial infections would no longer pose a significant threat to human health. Regrettably, modern pathogens have evolved multi-drug resistant strains, which are immune to the majority of available antibiotics[47] and hence the challenge of fighting bacterial infection has been reinstated, thus calling for further research. An additional adaptive mechanism employed by bacteria is the recently discovered prokaryotic immune system. This system is called CRISPR/Cas and allows bacterial cells to acquire immunity and adapt to invading bacteriophages[48,49].

The rules set up for this domain of life can mostly be considered universal. However, exceptions and variations to the rules are common[27,50,51] and may represent one of the central explanations for the overwhelming success of this group of life forms.

1.3.1 Overview of analyzed organisms

In the following, a brief description of the prokaryotic organisms investigated in this thesis will be provided. Two of them, *Escherichia coli* and *Salmonella enterica*, belong to the family of the Enterobacteriaceae while the other two, *Sinorhizobium meliloti* and *Agrobacterium tumefaciens*, are part of the Rhizobiaceae.

1.3.2 *Escherichia coli*

In human society, bacteria are mostly conceived in a negative context as they are usually associated with a lack of hygiene and disease. The same notion holds for the Gram-negative rod shaped bacterium *E. coli* (see Figure 3), which tends to appear in popular public media only when one of its pathotypes is causing malady or death. In 2011 for example, there was an epidemic in central Europe caused by sprouts contaminated with pathogenic *E. coli* bacteria[52,53], an occurrence that received significant media coverage. In the last decades, similar outbreaks have also been reported in the United States[54]. Indeed, various types of pathogenic and dangerous *E. coli* strains are known[55] and the investigation of their pathogenicity is an active and important field of research[56]. However, the fact that *E. coli* is a symbiont of paramount importance in gastrointestinal tracts of mammals is often overlooked.

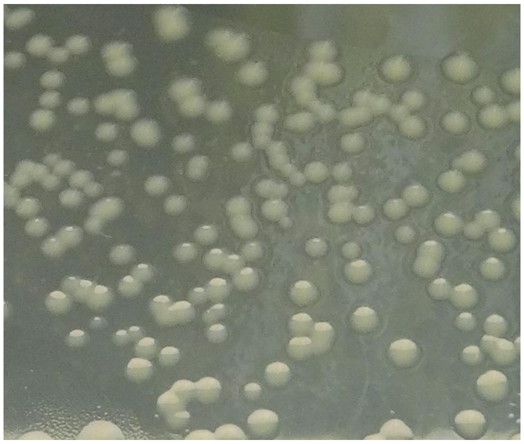


Figure 3. The photograph shows several *E. coli* colonies (beige spots) growing on a lysogeny broth[57] (LB)-agar plate. The image was taken and supplied courtesy of Dr. Stephan Klähn.

The non-pathogenic types of *E. coli* coexist with their host and the association is mutually beneficial[58,59]. Furthermore, *E. coli* has prevailed as an ideal model organism for the study of metabolic bacterial processes in general, and thus its 4.64 megabase (Mb) long chromosome was one of the early bacterial genomes to be completely sequenced[60]. To date, *E. coli* as a group represent one of the most in-depth studied and understood microorganisms. This is emphasized by the sheer volume of information in the EcoCyc database[61]. In EcoCyc, knowledge from literature is manually assembled and data are available for many *E. coli* genes. Research on this bacterium has led to major breakthroughs in science, including bacterial conjugation[62] and the operon model[63]. Maybe most important, however, is the discovery that *E. coli* can be used as a workhorse in modern molecular biological research and biotechnology where it may serve as a system to express heterologous genetic material, not originally stemming from *E. coli* itself[64–66]. In this capacity, *E. coli* has been used in the production of insulin[67,68], potential vaccines[69], and bio fuels[70].

The possibility of employing *E. coli* as a heterologous and genetically tractable system is also highly valuable when assaying RNA-RNA interactions in vivo. While an initial study showed that the protocol is viable for certain members of the *Gammaproteobacteria*[71,72], further studies validated the functionality of the procedure for *Neisseria meningitidis*[73] and cyanobacterial[74–76] RNA-RNA interactions.

For the work presented in this thesis, *E. coli* represents an invaluable resource since the majority of small RNA-target pairs have been identified here[77]. This resource was employed for the study presented in Chapter 4, to benchmark the quality of small RNA target prediction algorithms. Furthermore, the extensive annotation for *E. coli*[61] enables sound and highly informed interpretations of prediction results.

1.3.3 *Salmonella enterica*

It is common to frown upon un- or undercooked food and unboiled water or unwashed vegetables and fruit. This is due to the fact that food borne pathogens can occur in and on all of these types of nutrition and liquids. A widespread pathogen, responsible for the contamination of food and water, is the Gram-negative, rod shaped enterobacterium *S. enterica* (see Figure 4). World wide, over 2500 different types of *S. enterica* have been characterized[78].

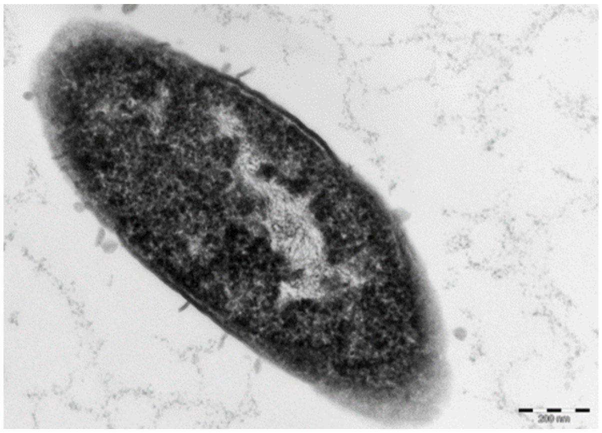


Figure 4. The image shows a single cell of *S. enterica* SL1344[79] taken with an electron microscope. The image was supplied courtesy of Prof. Dr. Kai Papenfort.

For the year 2000, 21.65 million *S. enterica* induced cases of typhoid fever were estimated. Of these, 216,510 were fatal[80]. A less disease and strain specific view estimates the annual number of *Salmonella* infections to be between 200 million and 1.3 billion[78]. Hence, *S. enterica* has developed to be a model organism for the study of bacterial pathogenesis in general. In the RNA field *S. enterica* has also become an important model for the in depth investigation of bacterial RNA-based regulators. Several RNAs well conserved throughout the *Enterobacteriaceae* have been most thoroughly characterized in *S. enterica*[81–88]. A widely investigated strain is *S. enterica* serovar Typhimurium LT2 that has a circular chromosome with 4.86 Mb and a single plasmid with 94 kilobases (Kb)[89].

Pathogenicity is strongly driven by the *Salmonella* pathogenicity islands (SPI). These SPIs are stretches of genetic information that encode factors important for a virulent lifestyle. Commonly, they are acquired by horizontal gene transfer[90]. The acquisition of such SPIs can rapidly accelerate the evolution of bacteria and can swiftly transform non-pathogenic bacteria to dangerous pathogens[91]. Two central SPIs in *S. enterica* are SPI-1 and SPI-2, which mediate invasion[92,93] of host cells and intracellular survival[94,95], respectively. Accordingly, a dual RNA sequencing study on human cells infected by *S. enterica* was able to show that SPI-1 genes connected to invasion were down regulated after successful host invasion, while SPI-2 genes important for survival within the host cell were up regulated[96]. Furthermore, the major post-transcriptional regulators Hfq and CsrA show pronounced binding in both SPI-1 and SPI-2, thus underlining the potential importance of these proteins for pathogenicity. These findings are outlined in more detail in Chapter 8.

1.3.4 *Sinorhizobium meliloti*

The majority of the earth's atmosphere is made up of molecular nitrogen (N_2)[97]. Even though it thus appears to be a plentiful resource, plants cannot directly tap into this vast reservoir because N_2 is a highly stable and inert molecule. Since nitrogen is a central component of biological macromolecules and many secondary metabolites, a limitation in its availability can severely impact a plant's growth[98]. To this end, leguminous plants have formed symbioses with rhizobial bacteria that are able to fix N_2 under application of the nitrogenase enzyme[99]. Although the majority of these unions indeed form between legumes and rhizobia, exceptions also exist[100,101]. In the symbiosis, plants internalize the bacteria into so called root nodules (see Figure 5). Here, the plant supplies energy and a microoxygenic environment for the bacteria who in turn focus on fixing N_2, which is made available to the plant as ammonium (NH_4^+). While recent research is focusing on unveiling the molecular details of these symbioses, the general knowledge about the benefit of growing leguminous plants as fertilizer in crop rotation has been known for millennia[102], even though the exact reasons for the added value must have been unclear to the farmers at the time. One of the most intensely researched organisms forming such a symbiosis is the Gram-negative bacterium *S. meliloti*, that has a genome size of 3.65 Mb. It also includes the two mega plasmids pSymA and pSymB that have lengths of 1.35 Mb and 1.68 Mb, respectively[103,104]. *S. meliloti* lives independently in soil or may specifically associate with certain plants[105] but not with others. Species of the genus *Medicago* are the common counterpart in the investigation of the symbiotic interaction of *S. meliloti* with leguminous plants[106].

For the successful interplay of *S. meliloti* with its plant host, close physical proximity and gene products from both organisms are required. Physical adjacency can be achieved through bacterial chemotaxis towards chemical attractants released by the plant[107].

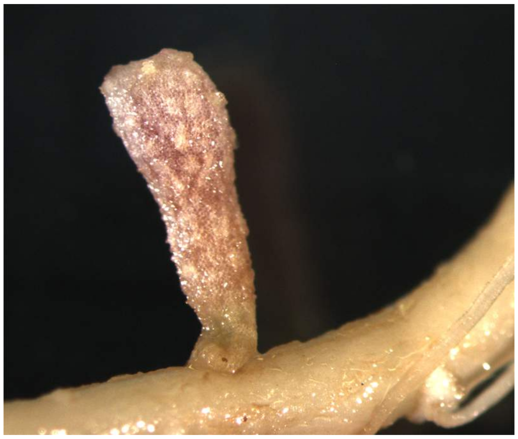


Figure 5. The photograph shows a *Medicago sativa* (Alfalfa) root nodule (red outgrowth on white root) formed through the symbiosis with the N_2-fixing rhizobium *S. meliloti*. The picture was taken and supplied courtesy of Dr. Marta Robledo.

When the partners are close, the infection begins with the plant sequestering flavonoids, which are sensed by the bacterial cells. Upon sensing the plant's flavonoids, the bacteria induce the production of Nod factors[108]. After this primary signal exchange, calcium levels in the host plant's root hairs start oscillating[109] and the root hairs curl up thus encasing bacterial cells[110]. In order to reach their final position in the plant's inner cortex, Nod factors and symbiotic exopolysaccharides from the bacteria work in concert with host gene products to form an infection thread that serves as a path for the bacterial cells through the outer cell layers of the plant's root[111].

Upon arrival at the final position within the plant's root cells, bacteria are enclosed by a host membrane, which gives rise to a structure that is called a symbiosome[106]. However, before the bacterial cells can shift their focus towards fixing N_2 they need to differentiate into a modified cell type referred to as a bacteroid. Several bacterial[112] and plant[113–115] derived genes are required for this differentiation.

A central property for an N_2 fixing environment is a strongly reduced oxygen content, because the nitrogenase enzyme can not properly function at high oxygen concentrations[116]. One of the factors that may facilitate the establishment of a low oxygen environment in root nodules are plant produced leghaemoglobin proteins[117]. These proteins give root nodules their characteristic red color (see Figure 5). Once the process of root nodulation is complete, *S. meliloti* has the ideal environment to commence N_2 fixation.

The importance of research conducted on such associations becomes apparent when considering that the soybean is also a legume, which forms root nodules together with the rhizobial species *Bradyrhizobium japonicum*. Soybean plays a central role in agriculture[118], and understanding the molecular details of its lifestyle is thus important. To this end, investigating *S. meliloti*'s association with plant symbionts can provide general insights into the nature of rhizobium-plant symbioses.

The investigation of bacterial small RNA regulators has also penetrated this field of research[119,120] and a small RNA (EcpR1) participating in the regulation of the cell cycle within *S. meliloti* has been characterized. The study highlighting these findings is presented in Chapter 7.

1.3.5 *Agrobacterium tumefaciens*

In 2001, two research groups concurrently published the full genome sequence of the Gram-negative, soil dwelling plant pathogen *A. tumefaciens*[121,122]. Unusually, it contains both a circular and a linear chromosome. These chromosomes have lengths of 2.84 Mb and 2.08 Mb, respectively. Importantly, *A. tumefaciens* also contains two plasmids, one of which is referred to as the t̲umor i̲nducing or Ti plasmid. This plasmid is vital for virulence since it contains the *vir*-genes that coordinate infection of plant tissue[123].

The infection has several central stages and commences with *A. tumefaciens* sensing and migrating towards plant attractants. While wounded plant tissue has been strongly implicated as a primary infection target, infection of unharmed plants has also been demonstrated[124]. The host recognition is mediated by sugars, low pH levels, low phosphate levels and plant derived phenols[125]. The bacterial recognition system is encoded by the *virA*, *virG* and *chvG* genes. Both *virA* and *virG* are located on the Ti plasmid while *chvG* is chromosomally encoded. Upon sensing attracting signals, the membrane bound VirA protein is phosphorylated and thereafter transfers the phosphate to the VirG protein, which resides in the cytoplasm. The ChvE protein, which is located in the bacterial periplasm, can sense a wide range of diverse sugar molecules and interact with the VirA/VirG system and sensitize it[126].

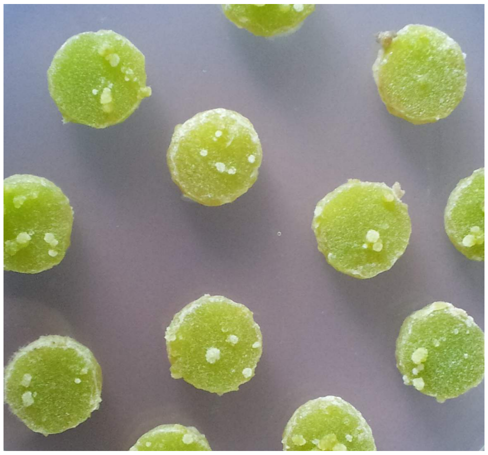


Figure 6. The image shows slices (one centimeter in diameter and four millimeters thick) of a potato (*Solanum tuberosum*) infected with *A. tumefaciens* bacteria on a water agar plate[127]. The white growths on the slices are tumors caused by the infection. The green color returns to the originally beige slices because they are subjected to a day-night cycle in the experimental setup, which causes the plant material to start producing chlorophyll. The picture was taken and supplied courtesy of AG Narberhaus.

The phosphorylated version of VirG is capable of activating the expression of further *vir* genes[128], which will finally cause the transfer of genetic material from the bacteria into the host plant's genome. This transfer DNA (T-DNA) is also encoded on the Ti plasmid. During infection, the T-DNA is excised from the Ti plasmid under application of the enzymes encoded at the *virD* locus. The single stranded, excised T-DNA is then coated and thus protected by VirE2 proteins[129]. Finally, the gene products of the *virB* and *virD4* locus form a type 4 secretion system, which provides a channel by which the T-DNA can pass the organismic barrier[130]. After arriving in the host, the T-DNA is incorporated into the plant genome and causes the production of opines, which can be metabolized by the bacteria[128]. Infected plant materials show obvious tumoral growths (see Figure 6).

Unlike *S. meliloti* (see Section 1.3.4), *A. tumefaciens* is pathogenic and has a wide range of potential hosts[131]. This property makes it not only hazardous to agricultural crops[132] but also useful as a molecular tool for the production of transgenic plants[133]. In line with this, transformations of important agricultural crop species have been attempted. Among others, successful transformations have been achieved with maize[134], barley[135], rice[136] and potato[137].

ATP-binding cassette (ABC) transporters play a vital role in the life cycle of prokaryotes. They aid in the transport of a wide range of molecules such as vitamins, sugars, amino acids and peptides[138]. The same holds true for *A. tumefaciens*, where several ABC transporters are controlled by a small RNA called ABC transporter regulator 1 (AbcR1)[139,140]. An initial study was able to identify three targets of AbcR1[139]. Chapter 5 contains a follow up study that vastly expands the regulon of AbcR1.

1.4 Prokaryotic *cis*- and *trans*-acting RNAs

Assuming an approximate setting in which an *E. coli* culture starts with a single bacterial cell and grows exponentially without encountering nutrient limitations, about 44 hours need to pass until the culture rivals the entire earth's weight[141]. Even though the planet is indeed densely populated by bacteria, they are not overgrowing it and this is due to the simple fact that nutrient availability is limiting their growth. To overcome such cues, prokaryotes have developed intricate regulatory networks that modify the cells' metabolism to the surrounding environment.

For a long time, proteins were considered as the central overseers of these networks. However, for more than a decade now, high throughput sequencing [14,142,143] and bioinformatics methods[144,145] have aided in the identification of a plethora of functional prokaryotic non-coding RNA (ncRNA) based regulators that perform tasks equally as important as protein-based regulators. For some prokaryotic organisms, the fraction of active transcriptional start sites (TSS) ascribed to ncRNAs may be 40% or higher[146].

The regulation imposed by these ncRNAs is often performed after transcription of their targets and is hence termed post-transcriptional regulation. The targets are commonly mRNAs and the regulative effect can have an influence on an mRNA's translation and/or stability. Several straight forward advantages of such RNA-based regulation are apparent. Firstly, post-transcriptional regulation constitutes an additional layer of control allowing a more sensitive adjustment of cellular metabolism. Also, a post-transcriptional regulator can encompass a regulon that is not directly connected at the transcriptional level[81]. Secondly, RNA-based regulators can be produced faster than protein-based regulators[147]. Furthermore, a lag in response time can be circumvented by post-transcriptional regulation, since direct repression of target transcripts prevents continuous production of protein products that are no longer appropriate for the changed environmental conditions. Low metabolic cost of RNA-based regulation has also been suggested as an advantage, but the impact of this has been challenged[25,148]. Finally, it has also become clear that RNA-based regulators can help avoiding negative effects of leaky transcriptional gene regulation[149]. Initial studies focused on *E. coli*[150–153], and *Salmonella*[154] but numerous further prokaryotes have since undergone similar characterization[120,142,146,155–163].

One class of RNA-based regulators are *cis*-antisense RNAs (asRNAs). The first functional asRNA mechanism was discovered on the bacterial ColE1 plasmid[164]. Since then, it has been established that asRNAs also occur chromosomally both in prokaryotes[165] and eukaryotes[166]. They are known as *cis*-asRNAs due to the fact that they are encoded on the opposite DNA strand but at the same locus as their target. *Cis*-asRNAs are widely distributed in prokaryotic organisms and have various functions such as transcriptional termination or translational control[165]. These functions are performed through the asRNAs' interactions with their targets via direct base pairing. An example of such an mRNA-asRNA interaction is the IsiA-IsrR pair from *Synechocystis* sp. PCC 6803. If both RNAs are present at the same time, an RNA duplex forms which is subsequently degraded[167].

Cis-acting RNA elements must, however, not necessarily be encoded on a different transcript. Riboswitches for instance are encoded on the same mRNA that they regulate[168,169]. Upon encountering specific metabolites, these riboswitches can change their structure and thus perform regulation. Depending on the riboswitch's structure, translation of an mRNA may be activated or deactivated. Similarly, RNA thermometers can sense temperature by changing their conformation in a temperature dependent manner and adjust whether an mRNA is translated or not by forming structures that render the ribosome binding site (RBS; also known as Shine-Dalgarno sequence[170]) accessible or inaccessible[171–173].

Further important members of the prokaryotic ncRNAs are *trans*-acting small RNAs (sRNAs)[23,25]. They can be considered as functional analogs of eukaryotic microRNAs (miRNAs)[174,175], and like their eukaryotic counterparts also use an initial seed base-pairing mechanism to mediate the interactions with their targets[176,177]. On an evolutionary scale, miRNAs and sRNAs appear to have originated more recently[25,178]. Prokaryotic sRNAs are between 50 and 500 nucleotides long and act by forming imperfect RNA-RNA duplexes with their target mRNAs. Many organisms employ the RNA chaperone Hfq to facilitate RNA-RNA interactions[179] (see Section 1.5.1 for a detailed description of this protein). Their targets are usually encoded at different loci, which is why this group of regulators is classified as *trans*-acting in contrast to the previously mentioned *cis*-acting elements.

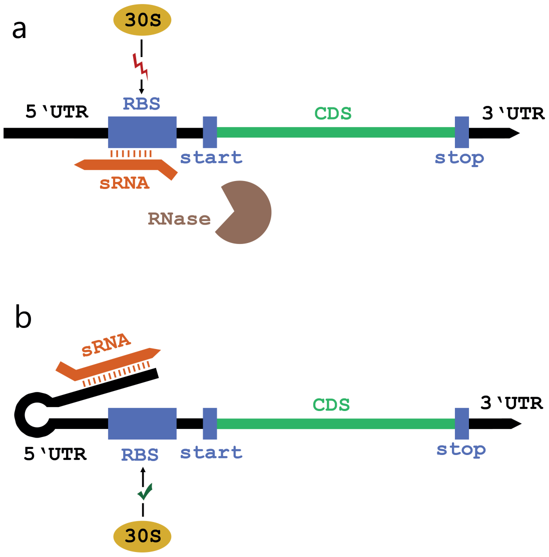


Figure 7. The figure shows two modes of action that have been reported for prokaryotic sRNAs (orange). Figure **a** depicts the negative regulation where an sRNA may bind to an mRNA's ribosome binding site (RBS, blue) and thus prevents binding of the small ribosomal subunit (30S). This inhibits translation initiation. In some cases, subsequent degradation of the RNA-RNA duplex by RNases (brown) follows in a process termed coupled degradation. Figure **b** shows how an sRNA can activate an mRNA target by breaking up intramolecular base pairs and rendering the RBS accessible for the small ribosomal subunit. The coding DNA sequences (CDS) are colored in green and the 5' and 3' untranslated regions (UTR) are colored in black. Start- and stop-codons are colored in blue.

However, sRNAs acting on *cis*-encoded targets as well as on *trans*-encoded targets have also been reported[180,181]. The majority of sRNAs studied to date are encoded in intergenic regions of prokaryotic genomes and are transcribed as individual units. This notion has since been extended by the results of a study investigating Hfq bound transcripts. The authors were able to show that 3'untranslated regions (3'UTR) can also be the source of functional *trans*-acting sRNAs[182].

Canonically, sRNAs base pair with the 5'untranslated region (5'UTR) of a target mRNA, but sRNA binding deep within coding DNA sequences (CDS) has also been observed[183–185]. Furthermore, archaeal sRNAs have been shown to exert their functions by binding to the 3'UTR of their target mRNAs[186]. The more common RNA-RNA interactions in the 5'UTR often occur in close proximity to the start codon or the RBS, which is situated upstream of the start codon (see Figure 7). Since both an accessible start codon and RBS are pivotal for translational initiation[11] their occlusion impedes efficient protein synthesis from a target mRNA. In fact, it has been shown that initiating ribosomes will cover the mRNA from nucleotide -21 (± 2) to +18 (± 1) with respect to the position of the start codon[187]. Clearly, blockage of this area by other factors such as sRNAs interferes with ribosome binding. This process is known as translational silencing. An sRNA driven inhibition of mRNA translation can also be achieved by blockage of ribosome standby sites[188,189], translational enhancers[81,190] or upstream open reading frames (ORF)[191]. Furthermore, negative regulation of targets may occur or be enhanced via sRNAs flagging mRNAs for degradation by RNases[192] similar to the previously discussed asRNA case of IsiA-IsrR[167] (see Figure 7a). The process in which paired RNAs are digested, is called coupled degradation[193].

Less common but also prevalent are cases in which sRNAs positively regulate target mRNAs[194–196]. Here, an mRNA may have formed an intramolecular structure that renders the components important for translational initiation inaccessible. The interaction with the sRNA then allows these structures to resolve and translation can commence (see Figure 7b). A positive effect of sRNAs binding to their targets can also be exercised by increasing a target RNA's stability[197,198], for instance by preventing RNase-based cleavage.

Bacterial sRNAs can furthermore act as cellular fishing-rods that sponge or trap target molecules by direct interaction. In this capacity, the CsrB RNA can titrate CsrA proteins and thus impose a regulatory effect[199–201]. The 6S RNA is also a well studied, RNA-based trap of the σ^{70} associated RNA polymerase holoenzyme[202]. By trapping σ^{70} linked RNA polymerases it represses the transcription from promoters that have a preference for this σ-factor, thus enhancing transcription from promoters favoring σ^{S}[25,203–205]. RNA-RNA interaction based target titration has also been shown[206–208].

While sRNAs are generally classified as non-coding transcripts, there are also reports on dual action sRNAs, which act by direct base pairing to target mRNAs and additionally contain an ORF, which is translated into a protein[209]. Recently, such RNAs that encode small proteins are gaining increased attention[210–213].

Prokaryotic *trans*-acting sRNAs are known to regulate many molecular pathways including virulence[214,215] and often represent regulatory hubs for specific pathways such as cellular amino acid metabolism[81,87,216,217], sugar metabolism[149,218,219], iron homeostasis[191,220] and photosynthesis[75]. Hence, their overall influence on cellular networking can be compared to that of globally acting transcription factors (TF), except that they act at the post-transcriptional level, instead of directly influencing whether RNA molecules are initially transcribed or not. However, since many sRNAs have been shown to post-transcriptionally regulate mRNAs that code for TFs they can also indirectly affect transcriptional regulation[77,84,87,183,191,221–223]. A recent study was furthermore able to establish a direct connection between sRNAs and transcriptional antitermination[224]. The authors showed that sRNAs are capable of stimulating full length mRNA synthesis by specifically binding the 5'UTRs of target mRNAs, which directly interferes with Rho-mediated[225] transcriptional termination.

In the following sections several specific enterobacterial sRNAs and their functions will be discussed. These sRNAs represent a subset of the input for the benchmark conducted in Chapter 4 in which the performance of the novel sRNA target prediction algorithm CopraRNA[77,226] is compared to other approaches.

1.4.1 FnrS

The transition from aerobic to anaerobic lifestyles requires extensive rewiring of active cellular production. For example, the products of genes such as *sodA*, which is responsible for the depletion of cellular levels of superoxide (O_2^-)[227], can be reduced in anoxia since a significant amount of superoxide is mostly generated when cells grow in oxygenic environments[228]. On the transcriptional side, this restructuring is in part performed by the fumarate nitrate reductase (FNR) regulator[229], which is able to directly sense the local oxygen concentration[230]. However, even though initial studies were able to link FNR to the restructuring[229,231,232], a regulation solely performed by FNR did not seem to explain all aspects of the bigger picture[233].

Light was shed into the dark by the discovery and characterization of the FNR-regulated sRNA (FnrS), which participates in reprogramming of cellular protein biosynthesis during the transition to anoxic conditions by exerting its function on the post-transcriptional level[233,234]. Indeed, as the name suggests, FnrS is transcriptionally activated by FNR and only abundant during anaerobic growth. This approximately 120 nucleotide long sRNA is well conserved within enterobacteria and thus can be found in species such as *E. coli*, *S. enterica* and *Yersinia pestis*[235]. FnrS was first discovered, and originally named RydA, in a comparative genomics study, where its expression could, however, not be confirmed[150], probably due to its anaerobic expression pattern. Later studies proved its existence[233] and direct association with the Hfq protein in vivo[236]. Two independent and simultaneous microarray and proteomics driven studies in *E. coli* described the first known direct FnrS targets[233,234] and thus helped in finding the missing link between the FNR protein and some of its indirect target RNAs.

1.4.2 GcvB

Amino acids play a central role for all living organisms as building blocks for proteins. Thus, a shortage of amino acids can lead to strongly impaired growth of bacterial cells[237]. Furthermore, amino acids can be utilized as nitrogen and carbon sources. However, when certain prokaryotic cells enter exponential growth in nutrient rich media, the uptake and production of amino acid molecules is repressed[190]. On the post-transcriptional level, this global regulation of amino acid metabolism is performed by the approximately 200 nucleotide long, Hfq dependent glycine cleavage B (GcvB) sRNA, which is widely distributed in gammaproteobacteria[235]. Originally identified in *E. coli* and shown to be regulated by the GcvA and GcvR proteins[238], the majority of RNA targets to date have been confirmed in *S. enterica*[81,87]. However, many of these GcvB target interactions are most likely conserved and there are even examples of targets that have been experimentally confirmed for both *E. coli* and *S. enterica.* These targets are the *sstT*[87,217] and *cycA*[87,216] mRNAs. On the other hand, certain interactions like GcvB-STM4351[81] appear to be more organism specific[77]. Early studies were able to identify an accessible GU-rich linker region 1 (R1) as an important domain for target repression. Using this R1 domain, GcvB represses its targets by either blocking Shine-Dalgarno or translational enhancer sequences[81,190]. Follow up studies extended the GcvB regulon to further mRNA targets in *S. enterica* and *E. coli*[87,190,239,240]. Among these targets are the transcription factor mRNAs encoded by the *lrp*[87], *phoP*[240] and *csgD*[239] genes. All studies combined make GcvB the sRNA with the currently largest known direct post-transcriptional network. Interestingly, GcvB itself can be repressed by other RNAs via an “anti-sRNA” pairing mechanism[206,207]. One of these anti-sRNAs is SroC, which is encoded between the *gltI* and *gltJ* genes and emerges as a stable RNA fragment when mRNA from the *gltIJKL* locus is degraded. SroC reduces the half-lives of GcvB molecules to under 2 minutes[241] in an RNase E dependent manner[206].

Finally, GcvB represents the archetype of an sRNA for which targets predictions are successful, as can be seen in the benchmark performance presented in Chapter 4.

1.4.3 RyhB

Iron is a vital cofactor for many proteins. For the previously mentioned transcriptional regulator FNR for instance (see Section 1.4.1), the ability to sense oxygen concentration strongly depends on an iron-sulfur cluster associated with the protein[230]. However, certain proteins are more important than others and cells encountering iron scarcity need to repress the production of less crucial iron-containing proteins to maintain appropriate levels of iron-binders that are essential for survival. On the other hand, excess iron can cause damage to the cell by aiding in the formation of reactive oxygen species[242]. Both situations serve to illustrate that cellular iron levels need to be tightly regulated to prevent toxic effects while also maintaining the functionality of systems requiring iron. Transcriptionally, the ferric uptake regulator (Fur)[243] performs regulation of at least 90 genes in *E. coli*[244]. Many Fur targets are repressed during iron availability, by iron bound Fur directly associating with so called Fur boxes in regulated promoters. Binding of Fur subsequently prevents transcription[245]. When iron becomes scarce, iron is released from Fur and Fur driven repression is lifted. Positive regulation by Fur was also observed, but due to the absence of Fur boxes in the promoters of these genes the activation was deemed indirect or at least non-canonical[246,247].

Indeed, a 2002 study in *E. coli* discovered that the Fur repressed, 90 nucleotide long sRNA RyhB is a post-transcriptional regulator of iron homeostasis and hence commenced to shed light into the thus far puzzling Fur network[248]. The authors of this study were able to show that RyhB negatively regulates iron related genes such as *fumA*[249] or *sodB*[250] during iron scarcity.

RyhB, which was originally found a year earlier[150] is well conserved within proteobacteria and requires Hfq for its activity[248]. In the fifteen years since its discovery numerous studies[191,193,220,251–262] have contributed to the ever expanding knowledge on the regulon of this sRNA and have shown that RyhB is not only capable of target repression but that it also directly and indirectly activates[195,220,254,261] targets that are important when iron needs to be scavenged from the surrounding environment. Of note, RyhB is also able to directly repress its own repressor, Fur, by targeting an upstream ORF[191].

Given the previously explained importance of tight iron availability regulation, it comes as no surprise that other organisms have evolved functional analogs of RyhB to regulate iron homeostasis. Among these are the PrrF RNAs in *Pseudomonas aeruginosa*[263], the FsrA RNA in *Bacillus subtilis*[264,265] and the IsaR1 RNA in *Synechocystis* sp. PCC 6803[159].

1.4.4 Spot42

Catabolite repression is a text book example for bacterial physiology and gene regulation. It is the umbrella term for the process in which cells regulate which carbon source to consume first if confronted with several different options. For example, the availability of glucose will repress the uptake of alternative sugars in *E. coli*, since glucose is favorable[266]. Central players in this circuitry are the small molecule cyclic adenosine monophosphate (cAMP), the cAMP receptor protein (CRP) and a roughly 110 nucleotide long sRNA called Spot42 (see Figure 8) which is transcribed from the *spf* (spot forty two) locus. CRP, in concert with bound cAMP, activates the expression of genes required for the uptake of several non-preferred carbon sources. Examples include *xylF* for the uptake of xylose[267] and *mglB* for the uptake of galactose[268].

Furthermore, transcription from the *spf* locus is repressed by CRP-cAMP[269]. However, since cAMP and CRP levels decrease in the presence of glucose[270,271], the activation of loci for the use of non-preferred carbon sources is abolished and the repression of the *spf* locus is lifted. Consequently, the Spot42 sRNA is amply transcribed in the presence of glucose[272]. For a long time the specific function of this widely conserved, Hfq dependent[218,235,273] sRNA remained nebulous, and initial attempts to characterize the transcript's role naturally suggested that it may encode a protein[274].

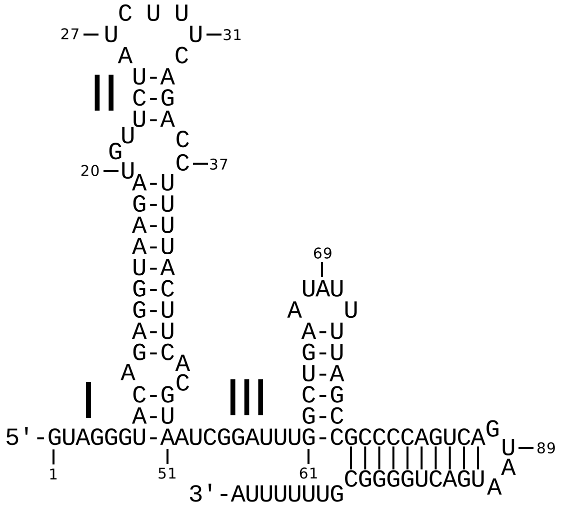


Figure 8. The figure shows the structure of the Spot42 sRNA of *E. coli* adapted from[149,218]. I, II and III highlight the regions of Spot42 that have been found to be important for target recognition. Positional numbering within the sRNA is also shown.

The first direct target of Spot42 identified was the *galK* mRNA. Here, Spot42 prevents ribosomes from binding to the Shine-Dalgarno sequence of the *galK* unit of the *galETKM* mRNA, causing discoordinate expression of this polycistronic mRNA[218]. This means that even though the proteins encoded on the *galETKM* operon are translated from the same mRNA they are not produced in equal quantities. However, *galK* remained the only known Spot42 target for nearly a decade. A microarray driven study in 2011 changed this and vastly expanded the known regulon of Spot42, thus cementing its role as a global player in the regulation of carbon metabolism[149]. The same study was also able to show that Spot42 employs three of its accessible regions (I, II and III) to bind target mRNAs (see Figure 8). The authors also found that Spot42 is involved in a coherent, multi-output feed-forward loop[275] together with the previously mentioned CRP, since many targets repressed by Spot42 are activated by CRP, and CRP in turn also represses Spot42 transcription. A follow up study employing computational target predictions was able to further extend the known Spot42 regulon[219].

Non-canonical target repression by the Spot-42 sRNA has also been reported. In an elegant experimental setup, which allowed a detailed mechanistic dissection of the interaction of Spot42 with the *sdhC* mRNA, Desnoyers and Massé[276] were able to show that the sRNA does not cause target repression itself, but it is rather involved in recruiting Hfq, the factor responsible for direct translational inhibition by blocking the translational initiation region.

Next to the *E. coli* variant, a Spot42 homolog that shares 84% identity with its *E. coli* counterpart has recently been investigated in *Aliivibrio salmonicida*[277]. An additional Spot42 target, the *mglB* mRNA, has been described in *S. enterica* (see Chapter 8).

1.5 RNA binding proteins

The impact of RNA molecules beyond the function of mere messaging has been soundly established in the scientific community[278]. However, due to the aim of classifying RNA-based functions and networks in more detail, some focus has again turned to specific proteins that directly interact with RNA molecules. These RNA binding proteins (RBP) directly bind target RNAs to exert their functions[279] and the abundance and diversity of proteins with RNA binding capacity may be higher than currently assumed[280,281]. The binding is mediated by RNA motifs to which the proteins specifically attach. These motifs can be sequence based, but may also contain structural RNA components[282,283]. RBPs have been functionally linked to processes such as alternative splicing[284], RNA-RNA interaction[179,285], and RNA stability[286].

The currently common wet-lab methods for detecting RNA-protein interactions on the genomic scale are referred to as crosslinking immunoprecipitation (CLIP) techniques[287–290]. Initially applied in studies on eukaryotes, these methods have now been extended to prokaryotes[201,207].

The following two sections provide a more detailed description of the bacterial RNA binding proteins Hfq and CsrA, for which CLIP-based genome wide binding maps were retrieved and analyzed in the work presented in Chapter 8.

1.5.1 Hfq

Historically, the Hfq protein was identified as a factor important for the infection of *E. coli* cells by the RNA bacteriophage Qβ, and its naming – host factor Qβ – also originates from the primary characterization[291]. In the many years since this initial study from 1968, the knowledge about the importance of the Hfq protein for *E. coli* itself has been ever increasing, and its role as a global player in post-transcriptional regulatory networks is now well-established[179,201,292], also beyond the context of *E. coli*.

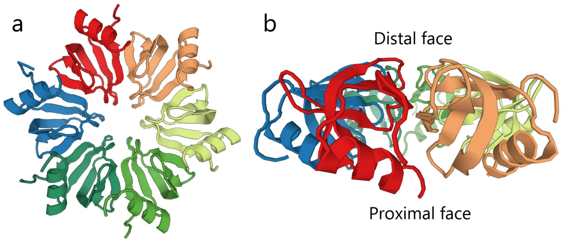


Figure 9. Two angles (**a**, **b**) of the three dimensional crystal structure of the *E. coli* Hfq protein[293] retrieved from the RCSB Protein Data Bank[294,295] and visualized with JavaScript Protein Viewer (PV)[296]. The proximal and the distal face are indicated in **b**. The colors highlight the individual domains of the protein.

Hence, bacteria that have lost their ability to produce Hfq proteins, can show defects in motility, secondary metabolite production and their response to stress. Virulent strains may even show loss of pathogenicity[297–300]. In contrast, the absence of Hfq does not show such obvious effects in all species. In *Staphylococcus aureus* for instance, *hfq* null mutants appeared to be no different from wild type[301].

Several crystal structures of Hfq from different organisms have been published[293,302–309], all confirming a homohexameric protein that is comparable to the shape of a dough nut (see Figure 9) and exhibits similarity to eukaryotic Sm[310] (for Stephanie Smith[311]) and Sm-like[312] RNA binding proteins[313,314].

Hfq is a vital RNA chaperone that mediates RNA-RNA interactions in many organisms[179], but interestingly shows little to no RNA affinity in others and may thus fulfill different purposes in distinct species[309,315,316]. Next to its role as an RNA-RNA interaction mediator, Hfq has been shown to have a function in the control of RNA stability. In this context, Hfq can protect RNA molecules from RNase degradation[179], but may conversely also enhance the RNase driven degradation of specific RNA species[193,317–319]. This function can also be exerted by Hfq dependent recruitment of the poly(A) polymerase enzyme to RNA targets for subsequent cleavage via an exoribonuclease[320].

For the Hfq variants with strong RNA affinity, researchers were initially confronted with a curious conundrum. Experiments analyzing the half lives of Hfq-RNA complexes were able to show that these complexes are stable and exhibit low dissociation rates[321–323]. Furthermore, the cellular pool of available Hfq proteins can limit sRNA activity[324]. However, Hfq has been shown to act as an in vivo facilitator of RNA-RNA interactions in many cases[197,219,325–328], and the cellular responses to stress, mediated by such RNA-RNA interactions are often fast[193,329]. This seemingly contradictory experimental evidence has since been resolved by the suggestion of an experimentally supported model in which RNA is actively cycled on Hfq[330]. In this model, different RNA molecules compete for the binding sites on the Hfq protein, consequently displacing each other and thus shifting the reaction times to the appropriate temporal frame that is important for the correct functionality of RNA-RNA interaction networks in vivo.

The three binding faces of Hfq are referred to as the distal, the proximal and the lateral face, which is also known as the rim (see Figure 9). The distal face has a preference for A-rich RNA sequences containing ARN (adenine, purine, any nuleotide) motifs[302], the proximal face preferably binds Us[307], and the lateral rim of the protein has been shown to interact with UA-rich RNA sequences[331,332]. Variants of the Hfq protein in different species can include unstructured regions of consecutive amino acids at the carboxyl ends of the protein[333–336]. While certain functional links have been established, the exact global functions of these diverse C-terminal tails in different species are still partly elusive[335,337–339].

Original studies aimed at finding RNA partners of the Hfq protein used co-immunoprecipitation (coIP)[182,236,340,341], which is, however, not able to detect the precise interaction position between RNA and Hfq on the transcript. More recently, the application of CLIP to pathogenic *E. coli*[207] and *S. enterica*[201] has circumvented the limitations of plain coIP and is able to detect Hfq binding sites on target RNAs at single nucleotide resolution (see Chapter 8).

1.5.2 CsrA

Like Hfq, the homodimeric carbon storage regulator A (CsrA) is an RBP of global importance for prokaryotic organisms[200,342,343]. CsrA is widespread in the prokaryotic phylum[344,345] and next to the enterobacterial variants[346,347], homologs of this protein can also be found in other genera such as *Erwinia*[348,349] and *Pseudomonas*[350,351] where they are referred to as repressor of stationary-phase metabolites (Rsm) proteins.

Similar to sRNAs, CsrA proteins canonically act as post-transcriptional regulators of gene expression by directly binding to target RNAs. Binding can occur in 5'UTRs, CDS and 3'UTRs[201,352]. CsrA binds RNAs by recognizing sequence and structure motifs in its targets. Specifically, it has been found that CsrA has strong affinity for GGA-motifs. The crucial role of the GGA-motif for CsrA target recognition has been confirmed by SELEX[353] (systematic evolution of ligands by exponential enrichment)[354] and genome wide CLIP experiments (see Chapter 8). The aforementioned studies were also able to establish a structural motif. Both found that the GGA-motifs were predominantly located within the loops of CsrA bound hairpin structures.

When bound to a target RNA, CsrA can exert positive and negative effects on the protein output of specific mRNAs. Just like sRNAs, CsrA commonly interferes with translation initiation. It can bind and occlude the Shine-Dalgarno sequence of its target and thus interfere with ribosome binding. Examples of such targets directly inhibited by CsrA are the *E. coli cstA*[342], *glgC*[355], *nhaR*[356], and *hfq*[357] mRNAs as well as the *Salmonella hilD* mRNA[358]. Interestingly, CsrA negatively autoregulates the translation of its own mRNA by the same mechanism[359]. Furthermore, CsrA can destabilize its target mRNAs as reported for the *pgaA* transcript in *E. coli*[360]. Not as common in the current literature but also studied and partially confirmed are examples of direct, positive CsrA mediated regulation[361–364].

The first investigations of the CsrA protein in *E. coli* showed that it has a role in the regulation of glycogen metabolism, which also led to its naming[346]. Since this initial report on CsrA, further functional roles in many different organisms have become apparent. Among these are regulation of pathogenesis in *Salmonella*[358], *Helicobacter pylori*[365] and *E. coli*[366], biofilm formation in *E. coli*[360], cell motility in *B. subtilis*[367], *E. coli*[361] and *Campylobacter jejuni*[368] and quorum sensing in the human pathogen *Vibrio cholerae*[369], to name a few.

The function of *E. coli* CsrA is negatively regulated by the sRNAs CsrB and CsrC, which both contain multiple CsrA binding sites (eighteen and nine, respectively), and are thus able to effectively titrate CsrA proteins. Hence, overexpression of these sRNAs can activate pathways which are usually down regulated by CsrA[199,370]. CsrB and CsrC are in turn negatively regulated by a protein called CsrD, which flags them for

RNase E mediated degradation[371]. Furthermore, a recent study has shown cAMP-CRP mediated transcriptional repression of *csrC* in *E. coli*[372].

Given its central role in pathogenicity[373], CsrA has been suggested as a potential target for antimicrobial agents[374].

1.6 RNA interactions

Analogous to DNA, RNA molecules also allow for base pairing when complementary nucleotides are present. While T pairs with A and G pairs with C in DNA, A pairs with U and G can pair with C or U in RNA. Further pairs with the Inosine nucleotide are possible in vivo[375–377], but they are not considered in the work presented here. An RNA's set of base pairs defines its secondary structure. This structural model is two dimensional. Extended complementarity leads to the formation of stacked base pairs and thus to stable RNA duplexes. These duplexes may form within a single molecule of RNA or between distinct RNAs. The former are referred to as intramolecular interactions. This type of interaction produces a structured RNA that is composed of a set of loops as illustrated in Figure 10. These two dimensional secondary structures alone do, however, not completely represent the biologically active three dimensional conformation of any given RNA, since they do not consider factors such as crossing base pairs (i.e. pseudoknots). Pseudoknots are often omitted in secondary structure prediction due to the increased computational cost attached to their inclusion. In fact, it has been shown that consideration of pseudoknots in RNA secondary structure prediction represents an NP-hard problem[378]. However, the validity of secondary structures without crossing base pairs for biological modeling is widely accepted and much easier to predict than the entire three dimensional conformation[379].

The first efficient algorithm to predict secondary structures of RNA molecules was introduced by Ruth Nussinov and colleagues, and performs a dynamic programming based maximization of base pairs in $O(n^3)$ time and $O(n^2)$ space complexity, where n represents the number of nucleotides in the RNA sequence[380].

Typically, this maximization strategy does not give rise to structures that appropriately represent the in vivo or in vitro active conformation of an RNA. To this end, energy-based approaches have been developed, which are more successful in predicting accurate secondary structures by utilizing the so called nearest neighbor energy model. This model, scores the energy contribution of a structural component based on its immediate context. Specifically, base pair stacks have a stabilizing effect and contribute negative energy scores to an RNA's secondary structure. Conversely, stretches of unpaired nucleotides within structured regions (e.g. hairpin loops, bulges) cause destabilization and thus contribute positive energy scores. In summary, the overall energy of an RNA structure is the sum of the energy contributions of its individual components. The energies for the different structural components have been measured experimentally[381–383]. The unit of measure for the energy values is kilocalories per mol (kcal/mol) where lower energies imply more stable structures. Given this background, an optimal RNA secondary structure can be predicted algorithmically with dynamic programming that minimizes the overall energy and thus gives rise to the minimum free energy (mfe) structure in $O(n^2)$ space and $O(n^3)$ time complexity[384].

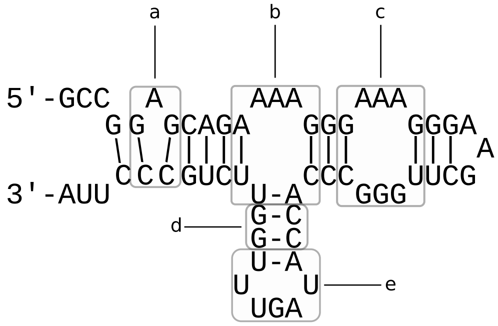


Figure 10. The figure shows the loop types an RNA's secondary structure can be composed of for an imaginary RNA sequence. Single examples of the possible loops are marked; **a** indicates a bulge-loop; **b** indicates a multi-loop; **c** pinpoints an interior-loop; **d** shows two stacked base pairs; **e** indicates a hairpin loop.

RNA folding based on free energy minimization has been implemented in programs such as RNAfold[385] and mfold / UNAFold[386,387].

One of the most intensely studied structured RNAs is transfer RNA (tRNA), that forms a characteristic clover leaf-like secondary structure[388,389]. The structure is often vital for an RNA's correct functionality, as visible in the tRNA clover leaf's accessible anticodon, which is important for correct codon recognition on the mRNA during translation. Prokaryotic sRNAs also fold into secondary structures, which render specific portions of the molecule accessible. These unfolded regions can, like for tRNA, be important for the correct recognition of target RNAs (see accessible regions I, II and III of the Spot42 sRNA in Figure 8).

Such intermolecular or RNA-RNA interactions will be the focus of the following paragraphs, where an outline of central points and different types of algorithms that attempt in silico RNA-RNA interaction prediction will be given.

1.6.1 Central aspects in RNA-RNA interactions

As outlined earlier (see Section 1.2), two strands of complementary DNA can form anti-parallel duplexes, which run from 5' to 3' on one strand and from 3' to 5' on the other. The same logic holds for RNA molecules and hence the assumption that base pair complementarity is the only factor ruling whether RNA molecules can interact seems initially reasonable. However, while such complementarity is the foundation for every RNA-RNA interaction it is not the only vital component. One of the additional components are the previously mentioned intramolecular structures that are established before the distinct interacting RNAs come into close contact with each other. A simple example employing the hypothetical RNA sequence 5'-CCCCCCCCCCGGGGGGGGGG-3' [a] readily serves to outline how neglecting or considering intramolecular base pairing can lead to strongly differing duplex predictions. Without

[a] By definition, RNA sequences are written in 5' to 3' direction.

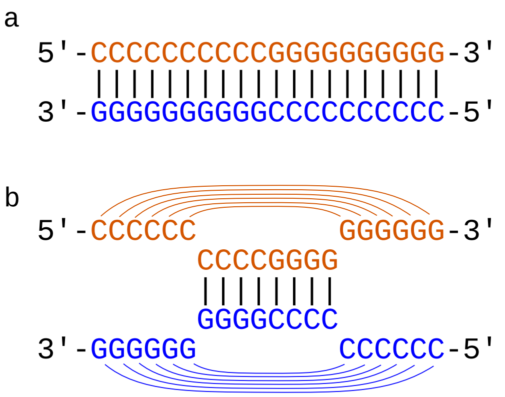


Figure 11. The figure shows RNA-RNA duplexes produced by two types of predictors for a pair of RNAs with the sequence 5'-CCCCCCCCCCGGGGGGGGGG-3'; **a** represents the duplex prediction returned by the RNAhybrid webserver[390]; **b** shows the duplex as predicted by the IntaRNA webserver[226] extended by potentially present intramolecular interactions. The black pipe characters indicate intermolecular base pairs, while the colored brackets represent intramolecular interactions. The two respective RNA molecules are colored orange and blue.

consideration of intramolecular folding a perfect duplex can be predicted. This scenario is depicted in Figure 11a. However, the assumption that an RNA molecule of this sequence does not fold is highly implausible. It is more likely to assume that a small hairpin structure with a stable G-C stem (arcs in Figure 11b) and a small accessible loop has developed before two RNAs of this type start interacting. In line with this, a more realistic model will only predict the accessible nucleotides to interact, as shown in Figure 11b. Algorithms for RNA-RNA interaction prediction are divided into two major classes based on whether they account or do not account for intramolecular base pairs. In vivo, additional factors such as temperature, chemical environment, small ligands[168,169] and RNA binding proteins[391,392] may influence an RNA's structure. Chapter 8 outlines how RNA binding protein data for the enterobacterial Hfq protein can be utilized to improve genome wide target prediction for bacterial sRNAs.

1.6.2 RNA-RNA interaction prediction without accessibility consideration

Many of the early approaches that attempted to predict RNA-RNA duplexes, neglected intramolecular base pairs. However, even in this class of predictors a line needs to be drawn between purely sequence-based and structure-based approaches. The sequence-based models search solely for complementarity. To find stretches of consecutive complementary bases the basic local alignment search tool (BLAST)[393] algorithm is a good initial approximator. BLAST is, however, limited to the standard Watson-Crick base pairs between G-C and A-U and will not consider the G-U wobble pair. This limitation is circumvented by the GUUGle algorithm[394]. Since these solutions basically represent local alignments they naturally inherit the possibility to asses the statistical significance of a predicted interaction[395]. Yet, the lack of a thermodynamic energy model limits the applicability of these solutions when looking for a biologically relevant RNA-RNA duplex. Nevertheless, sequence-based approaches can be useful by reducing the search space for more complex algorithms. TargetRNA is an algorithm that includes both a sequence-based and an energy-based model[396]. The sequence-based model applies a scoring scheme that is comparable to that used by Smith and Waterman for local alignments[397]. TargetRNA's sequence-based scoring scheme furthermore favors Watson-Crick pairs and penalizes loops within the predicted interaction site. In the energy-based model of TargetRNA the scoring attempts to minimize the free energy of the RNA-RNA interaction. RNAhybrid[390,398], an approach that was developed to predict eukaryotic miRNA targets and is also applied in duplex prediction for bacterial sRNAs[399], also uses energy-based scoring. The prediction strategy follows the same rules that have been established for intramolecular structure prediction (see Section 1.6) and the basic components that need to be considered are also the different loop types (see Figure 10). Likewise, energy is measured in kcal/mol and the individual energy contributions stem from the same database[381–383]. Also, lower interaction energies reflect more stable RNA-RNA interactions. Long interior loops can significantly increase the computational complexity for RNA-RNA interaction predictions and hence TargetRNA and RNAhybrid both restrict the maximum length for these structural elements. RNAplex[400] also uses energy-based scoring for RNA-RNA interaction prediction but differs from the two previously mentioned approaches in its handling of long internal loops and bulges. Instead of setting a fixed threshold for the length of internal loops, it employs an affine function to score them. A central limitation of approaches not considering the intramolecular structures of the interaction partners are disproportionately long duplex predictions. RNAplex attempts to tackle this problem by introducing a penalty for every nucleotide in the interaction.

In summary, the approaches neglecting intramolecular RNA structures represent a sensible initial approximation of in vivo RNA-RNA interactions. Especially the variants that include a thermodynamic model can certainly be regarded as a step towards a more realistic scenario, since they allow the consideration of temperature. In vivo, the temperature of the environment can play an important role for cellular dynamics and molecular structure. This is apparent for RNA thermometers[171–173], biomembranes[401] and for proteins[402]. RNA-RNA interactions follow the same principle and duplex predictions may have to be performed at high temperatures for organisms like thermophiles, which live in hot environments[403], to return meaningful results.

1.6.3 RNA-RNA interaction prediction considering intramolecular base pairing

Even though the previously mentioned RNAplex algorithm includes a penalty for every nucleotide in the interaction, and thus attempts to prevent overestimation of RNA-RNA duplex lengths, it does not specifically address intramolecular structures of the putatively interacting RNAs. The example in Figure 12 shows how neglecting the accessibility of RNAs within the predictive

```
a
argT 5'-…GUC                 CACGUAAAACAUAAG…-3'
            UGCAAACAC ACAACGC
            ||||||||| |||||||
            ACGUUUGUG UGUUGUG
GcvB 3'-…UUA         U       UUGUAGUGUUGGCAU…-5'
b
argT 5'-…GUC                 C         A A
            UGCAAACAC ACAACGC ACGU A AA C UAAG…-3'
            ||||||||| ||||||| |||| | || | ||||
            ACGUUUGUG UGUUGUG UGUA U UU G AUUC…-5'
GcvB 3'-…UUA         U       U    G G  G C
c
argT 5'-…GUCU                C  GUAAAACAUAAG…-3'
             GCAAACAC ACAACGC AC
             |||||||| ||||||| ||
             CGUUUGUG UGUUGUG UG
GcvB 3'-…UUAA        U       U  UAGUGUUGGCAU…-5'
```

Figure 12. The figure shows three conceivable RNA-RNA duplexes for the GcvB sRNA (blue) and its target argT/STM2355 (orange); **a** shows the experimentally confirmed in vivo interaction[81]; **b** depicts a putative duplex that may have been predicted without considering the accessibility of GcvB and argT; **c** shows the interaction as predicted by the accessibility-based IntaRNA webserver[226]. The black pipe characters indicate intermolecular base pairs. The figure is in part adapted from Backofen, 2014[404].

model can lead to overly long and consequently wrong duplex predictions when compared to experimental data from the wet-lab. For this reason, state of the art predictions consider how accessible the putatively interacting RNA regions are, and penalize interaction predictions between regions that appear to be entangled in intramolecular structures[144].

The following sections cover two common accessibility-based approaches. These are based either on concatenation or specific calculation of the accessibilities for the supposedly interacting RNAs.

1.6.4 Concatenation-based approaches

A plausible solution towards consideration of intramolecular structures within RNA-RNA interaction prediction are concatenation approaches. These algorithms predict interactions by first artificially fusing the RNAs to a single sequence under application of a spacer, which is referred to as the linker. After the RNA sequences have been concatenated RNA folding in a single molecule manner can be applied. From this approach it is already intuitively clear that intramolecular structures, which lead to a more stable secondary structure of the concatemer will be favored over intermolecular base pairs and hence the accessibility is embedded in the model. On the other hand, base pairs that represent stable intermolecular structures can be considered at the same time, thus allowing prediction of putative RNA-RNA duplexes.

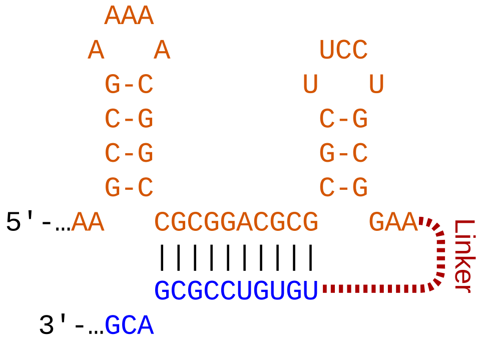


Figure 13. The figure shows a potential RNA-RNA interaction predicted by a concatenation-based approach. Pipe and dash characters indicate inter- and intramolecular base pairs. The two RNAs are colored in orange and blue, respectively. The linker is depicted in red.

RNAcofold[405], NUPACK[406] and PairFold[407] are examples of such concatenation-based predictors. Given that the linker is artificially introduced these algorithms need to extend traditional RNA folding to handle its presence. They do this by remembering its location and adjusting the energy calculation for the recursive cases that include the linker sequence. An example of this can be seen in Figure 13. Here the linker is part of an energetically unfavorable bulge loop. However, this bulge is artificial and should rather be treated as a dangling end. Hence, the penalty imposed for this structural element must be appropriately adjusted.

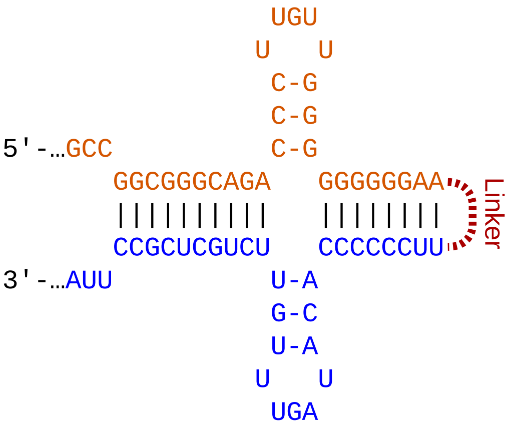


Figure 14. The figure shows a potential interaction predicted by a concatenation-based approach that includes a multi-loop in the RNA-RNA interaction region. Pipe and dash characters indicate inter- and intramolecular base pairs. The two RNAs are colored in orange and blue, respectively. The linker is depicted in red.

A major upside of concatenation approaches is that they natively inherit the properties of individual RNA folding. Hence, the partition function and base pair probabilities may also be calculated by application of McCaskill's algorithm[408]. Additionally, RNA-RNA interactions that form a multi-loop structure can be captured in the duplex predictions (see Figure 14). Since traditional RNA folding does, however, not allow pseudoknots, concatenation approaches are unable to predict interactions between RNAs that involve kissing hairpins (see Figure 15). This is a significant pitfall because such kissing hairpin interactions are known to be functional in vivo. The enterobacterial sRNA OxyS for instance targets the *fhlA* mRNA via kissing hairpins[409,410]. The interaction of the enterobacterial sRNA RyhB with the *sodB* mRNA[251] also represents a pseudoknot in the concatenation model and the *S. aureus* RNAIII similarly uses a loop-loop interaction mode to bind its targets[411,412].

1.6.5 Accessibility-based approaches

The inherent inability of concatenation-based approaches to predict pseudoknots is a central limitation, which can be circumvented by accessibility-based approaches such as IntaRNA[176] and RNAup[413]. Both algorithms are able to predict structures like the commonly occurring kissing hairpins (see Figure 15). Instead of merging the RNA sequences and folding the concatemer, accessibility-based solutions address the structuredness of the potentially interacting RNAs individually. Predicted interactions between regions that exhibit pronounced secondary structures are hence penalized, which in turn favors interactions between regions such as accessible

RNA hairpin loops. Like for the approaches not considering the accessibility of the predicted duplex, accessibility-based solutions also compute the hybridization energy (E^{hyb}) for the duplex. However, they extend this model by computing the unfolding energies (ED^{R^1}, ED^{R^2}) for the interacting RNAs (R^1, R^2). The unfolding energies represent the energy necessary to transform intramolecular double stranded regions into single stranded ones. The hybrid energy for the duplex can be calculated under application of the previously mentioned strategy employed in RNAhybrid. The unfolding energies can be computed using a partition function based approach. Specifically, the partition function (Z_S) for all potential topologies (S) of a given RNA sequence is defined as:

$$Z_S = \sum_{P \in S} e^{-\frac{E(P)}{RT}}. \qquad (1)$$

Here, $E(P)$ represents the free energy of a specific structure $P \in S$ that an RNA can give rise to. Furthermore, the gas constant is represented by R while T denotes the temperature. From this, the ensemble energy (E^{ens}) of all potential structures S for a given RNA can be computed with:

$$E^{ens}(S) = -RT\ ln\ (\ Z_S\ ). \qquad (2)$$

Based on this, the ED^R value for a specific RNA (R), which represents the energy needed to make a range of bases starting at position i and ending at position j unpaired, can be calculated by subtracting the ensemble energy E^{ens} of all potential structures S from the ensemble energy of all structures with (i, j) unpaired $S^{unpaired}_{(i,j)}$.

$$ED^R_{(i,j)} = E^{ens}(S^{unpaired}_{(i,j)}) - E^{ens}(S) \qquad (3)$$

Both the energy of the entire ensemble and the energy of subsets of the ensemble can be computed with McCaskill's algorithm[408]. Since the resulting ED^R value is positive it can be considered as a penalty because lower energies represent more stable structures. Given the ED values (ED^{R^1}, ED^{R^2}) for the two potentially interacting RNAs (R^1, R^2) the hybridization energy (E^{hyb}) can be adjusted to account for intramolecular structures by adding the ED values to the hybridization energy. This yields the final extended hybridization energy (E^{ext}) for a predicted interaction between the positions i, j on R^1 and k, l on R^2.

$$E^{ext}_{(i,j,k,l)} = E^{hyb}_{(i,j,k,l)} + ED^{R^1}_{(i,j)} + ED^{R^2}_{(k,l)} \qquad (4)$$

IntaRNA furthermore extends this solution by enforcing a seed constraint. Commonly, the seed is a stretch of six to eight consecutive complementary base pairs. This extension enhances the predictive performance[176] and is biologically warranted for eukaryotic miRNAs[414] and prokaryotic sRNAs[25]. Even though accessibility-based solutions are currently the most successful approaches towards in silico prediction of in vivo RNA-RNA hybrids, they also have their limitations. Firstly, and in contrast to concatenation-based approaches, current implementations of accessibility-based algorithms are not able to predict interactions that involve a multi-loop structure.

```
              AUA   GUA
5'-…AAC      G   C-G   A       AAG…-3'
       CCGGGG    C-G    GCCGGGG
       ||||||    C-G    |||||||
       GGCCUU    C-G    CGGCUUC
3'-…AAA      A   C-G   A       AAG…-5'
              AGA   AAA
```

Figure 15. The figure shows two hypothetical RNA molecules forming an RNA-RNA duplex via loop-loop interaction. Pipe and dash characters indicate inter- and intramolecular base pairs. The two RNAs are colored in orange and blue, respectively.

Figure 14 shows an example of such a case. For two interacting RNAs of this type it may be possible to form such a duplex; however, accessibility-based implementations are only able to predict the duplexes independently. A non-artificial example of an RNA that may be able to act by such a pairing mechanism is the sRNA Spot42 which could use both of its unstructured regions I and III simultaneously (see Figure 8) to pair with its targets[218]. The second unpredictable interaction types are double kissing hairpins such as the previously mentioned interaction between OxyS and its target encoded by the *fhlA* gene[409,410]. General joint structure prediction approaches such as IRIS[415] and biRNA[416] attempt to predict these more complex interactions.

1.6.6 Comparative RNA-RNA interaction prediction

An issue that all RNA-RNA interaction prediction methods still have in common is their high false positive rate when attempting to predict sRNA targets on the genomic scale[417]. Specifically, this means most predictive approaches tend to fail in identifying an sRNA's correct regulon when the entire pool of mRNAs is considered in the target prediction. *S. enterica* for instance encodes over 4500 protein coding genes, which must all be considered within the target prediction[89]. One reason for the high false positive rate is the lack of knowledge about the in vivo setting. This can lead to somewhat artificial predictions based on an inferred in vitro system that only considers the two putatively interacting RNAs. Even though the assumptions made for the predictive system are sound, neglecting additional factors can lead to spurious results. An example of such a scenario is depicted in Figure 16. Here, the accessibility of an RNA molecule is compared with and without cofactors that bind to the RNA. Proteins, other RNAs and small ligands are conceivable cofactors. Without cofactors the blue RNA stretch between x and y appears to be highly accessible, while the green stretch between x' and y' is entangled within an intramolecular stem. Hence, an accessibility-based predictor will favor interactions with other RNAs that employ the blue region. However, when the cofactors are bound, the second stem is unfolded and the green region is unstructured. Furthermore, the blue region is masked by one of the cofactors, which makes it inaccessible. While this example may be extreme, it clearly outlines the challenges that prediction algorithms are confronted with.

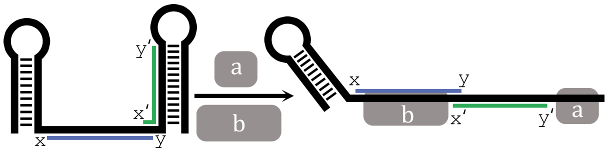


Figure 16. The figure shows two possible secondary structures of an RNA molecule (black). The left side depicts the topology when no cofactors a and b (gray) are involved, while the right inside shows the structure with bound cofactors. The positions x, y (blue) and x', y' (green) pinpoint specific regions of the structure for which the accessibility significantly changes when the RNA is bound by the two cofactors.

A popular remedy for high false positive rates in predictive computational biology is the use of comparative systems[398,418–421]. The rationale behind this approach is that conservation can be an indicator of functionality. In essence, such comparative approaches extend the predictions beyond the single organism context and assess whether the predictions hold for related species. Examples of comparative RNA-RNA interaction prediction algorithms are ripalign[422] and PETcofold[423]. Both predict the interactions based on alignments of the interaction partners. The original description of the RNAhybrid[398] algorithm also introduced a comparative scoring for miRNA target prediction. For this, the first step consists of transforming the duplex energies for the individual target predictions into p-values. This is important because duplex energies are not comparable when investigating distinct organisms. This is true for both eukaryotes and prokaryotes. An example how strongly the distributions of duplex energies can differ between organisms is shown in Figure 17. Here, the densities of duplex energies for whole genome target predictions of the GcvB sRNA (see Section 1.4.2) are compared across three different species. Clearly, all three organisms exhibit different distributions, which serves to demonstrate that duplex energies are not an appropriate measure for comparing RNA-RNA interactions in distinct organisms. Reasons for the heterogeneity in the distributions are differing GC-contents and dinucleotide frequencies in individual organisms.

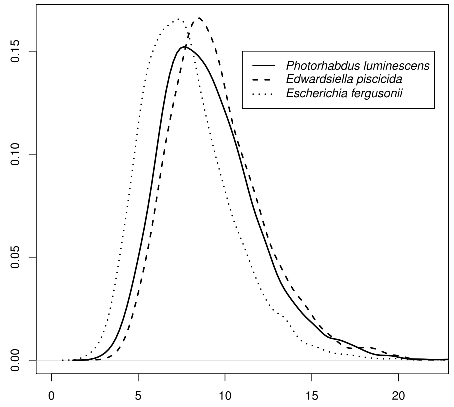


Figure 17. The plot shows the densities (y-axis) of the absolute IntaRNA[176] energy scores (x-axis) for GcvB sRNA (see Section 1.4.2) whole genome target predictions in the organisms *Photorhabdus luminescens* (NC_005126)[424], *Edwardsiella piscicida* (NC_020796)[425] and *Escherichia fergusonii* (NC_011740, NC_011743). The densities and the plot were calculated and produced using the density, plot and lines functions in the R statistics software[426].

The p-values can be obtained from raw scores by fitting an extreme value distribution to a background of energy scores, which can be produced by performing duplex predictions on dinucleotide shuffled (i.e. randomized) RNA sequences. The p-value itself represents the probability of returning a prediction of a specific quality or better given the background model. In other words, it contains information on how likely it is to get a certain prediction or better purely by chance. Since p-values are uniformly distributed between 0 and 1 p-value distributions for different organisms are comparable, while extreme value distributions of energy scores are not. Hence, transformation of raw energy scores to p-values represents a normalization. Given the now comparable individual p-values an appropriate combination strategy needs to be employed. One solution towards p-value combination was suggested by Ronald Aylmer Fisher[427].

$$X^2_{2k} \sim -2\sum_{i=1}^{k} ln(p_i) \qquad (5)$$

Here, k independent p-values (p_i) are combined and the resulting joined p-value can be retrieved from a chi-squared distribution with $2k$ degrees of freedom (X^2_{2k}). However, the individual p-values need to be independent and in the given case, where related species are compared, complete statistical autonomy is intuitively not satisfied. The reason for this is that all species are descended from a common ancestor, which automatically implicates a certain degree of mutual dependence.

Thus, the comparative solution in RNAhybrid[398] rather suggest a strategy for p-value combination that joins p-values of homologous putative targets to a joint p-value (p_{joint}) by selecting the highest p-value (p) in the set and raising it to the power of k, which is the number of species used in the comparative prediction.

$$p_{joint} = (max\{p_1, \dots, p_k\})^k \qquad (6)$$

As previously stated, the individual p-values cannot be assumed to be independent, which is why the authors of RNAhybrid continue to point out that the effective number of organisms (k_{eff}) needs to be calculated. The value for k_{eff} can be estimated between 1 and k.

$$1 \leq k_{eff} \leq k \qquad (7)$$

Specifically, k_{eff} can be assessed by dinucleotide shuffling the homologous miRNA sequences and performing subsequent duplex predictions for homologous targets. This gives rise to a background distribution of duplex energies, from which p-values can be derived. These p-values can be joined according to equation 6 by testing several *k'* values for *k*. The range for *k'* is between 1 and *k*. The *k'* that returns the straightest line in the empirical cumulative density function of the joint p-values is then chosen as k_{eff}. A smaller dependence infers bigger k_{eff} values and vice versa. A k_{eff} value of 1 would thus mean there is no benefit in performing a comparative prediction.

One pitfall of this and other comparative RNA-RNA interaction prediction approaches is that they tend to be overly restrictive. Specifically, the idea presented by RNAhybrid[398] only considers the largest (i.e. the worst) p-value for the calculation of p_{joint}. This means, only a single organism in the set of homologous targets needs to return a poor individual prediction to degrade the joint prediction for the entire cluster. While this strategy certainly reduces the number of false positive predictions it will also produce false negatives alongside. A less restrictive approach to genome wide comparative RNA-RNA interaction prediction, the CopraRNA algorithm, is presented in Chapter 4. Importantly, CopraRNA employs a p-value combination strategy introduced by Joachim Hartung that allows not only for the correction of individual p-value dependence (ρ) but also includes the possibility to assign individual weights (λ) to the single p-values during combination[428]. This is important because the organisms participating in a comparative prediction cannot be assumed to be evolutionarily equidistant. For this reason subgroups of organisms in the prediction that are more strongly related than others need to receive lower individual weights. For the combination the n individual p-values are first transformed to probits (t) and then combined with:

$$N(0,1) \sim \frac{\sum_{i=1}^{n} \lambda_i t_i}{\sqrt{(1-\rho)\sum_{i=1}^{n} \lambda_i^2 + \rho(\sum_{i=1}^{n} \lambda_i)^2}} \qquad (8)$$

The final combined p-value can then be derived from a standard normal distribution with a mean of 0 and a variance of 1.

1.7 Differential expression analysis

Instead of only performing in silico predictions scientists can also investigate the functional roles of cellular components or influence of external factors by performing experiments in the wet-lab. In classic genetics researchers investigated phenotypic differences between individuals and how these are passed on to the next generations. Modern genetics extend the traditional methods in order to uncover the molecular basis for the observed variations. For this, researchers often attempt to switch off a gene of interest[429]. Conversely, a gene of interest may also be overexpressed to unveil its function[430]. Resultant mutants are then compared to non-mutants, which are referred to as wild types (WT). The detailed molecular differences between mutants and WT can be investigated by sequencing and analyzing their transcriptomes[13,14]. The comparison of sequencing libraries is made possible by differential expression (DE) analysis methods. These techniques search for statistically significant, differentially abundant features in the transcriptomes[431,432].

Before the DE analysis can commence, the RNA-seq reads – that are nucleic acid fragments returned by the sequencer – need to be aligned to a reference. This task can be performed by so called mapping tools[433–436]. Then, because DE analysis techniques require count tables as input, the mapped reads need to be assigned to genomic features such as annotated mRNAs or ncRNAs. Methods performing these assignments count how many reads overlap with the features the researcher is interested in examining[437–439].

It is conceivable that two NGS libraries will not yield exactly the same number of reads[440]. This also influences the abundance of counts per feature in different libraries. In an extreme example one may assume that the sequencing depth in two libraries of exactly the same type differs by a factor of three or more. Without correcting for library specific read count abundance one might conclude that all features in the two libraries are differentially expressed for a biological reason. However, the true explanation for the difference in abundance is purely technical. For this reason read counts per feature need to be appropriately adjusted for library specific size differences prior to DE analysis. One of the original approaches used to normalize RNA-seq data introduced reads per kilobase per million (RPKM), which normalize counts for a given feature based both on sequencing depth and feature length[13]. RPKM normalization of counts for one library achieves this by first dividing the sum of all counts for the library by one million. All individual counts are then divided by this scaling factor followed by a further division by a feature's length in Kb. A variant of RPKM normalization is fragments per kilobase per million (FPKM) normalization[441], which adjusts RPKM to appropriately deal with paired-end sequencing data[b]. An inherent issue with this type of normalization is that changes in highly expressed genes will affect the normalization of counts for lowly expressed genes inappropriately, which can lead to faulty conclusions[432]. For this reason, more recent approaches such as DESeq[442,443] or edgeR[444] have introduced more appropriate normalization techniques that are based on the assumption that the majority of genes are not differentially expressed.

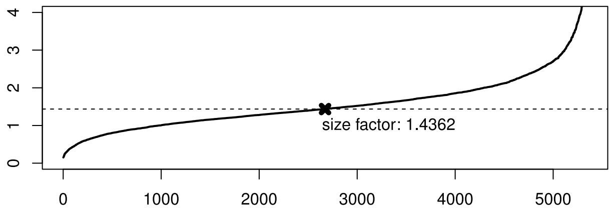


Figure 18. The plot shows a graphical representation of how DESeq selects the library specific size factor after the list of pseudo size factors has been computed (see text). The x-axis represents the indices of the pseudo size factors while the y-axis shows the magnitude of size factors. The data for the plot stems from the Hfq datasets analyzed in Chapter 8. The specific size factor that is indicated by the dashed line and the cross is the size factor calculated to adjust the counts for the first crosslinked Hfq library. The plot was generated employing the plot, points, abline and text functions in the R statistics software[426].

DESeq calculates a size factor to correct individual read counts by first computing a geometric mean of raw read counts for each feature over all libraries in the analysis. Then, it divides the raw counts for each feature by the geometric mean calculated for this feature. For every library this gives rise to a list of pseudo size factors that is of the same length as the list of features itself. Then, for each library individually, the median

[b] Paired-end sequencing is a method that can produce two reads per sequenced nucleic acid fragment, while single-end sequencing will only produce one[4].

of its pseudo size factors is selected as actual size factor. This is graphically presented in Figure 18. The sigmoid curve shows the pseudo size factors and the median is indicated by the dashed line and cross. The adjustment of raw read counts is performed by dividing each raw count in a specific library by that library's size factor. Overall, this procedure eliminates the strong influence that outliers can have in more traditional methods.

Currently, leading methods for DE analysis model the count data with negative binomial (NB) distributions instead of Poisson distributions, which were employed in earlier analyses of DE[445]. This is owed to the fact that NB distributions include the two variable parameters mean and variance to model count data from biological replicates. They are thus more accurate than Poisson distributions for which the variance equals the mean, which can cause an underestimation of the true variance (i.e. overdispersion)[442,443]. The DESeq2 algorithm also uses a NB model for which it firstly defines the parameters of the distribution and then computes p-values under application of a Wald test[443]. For the work presented in this thesis, DESeq2 was employed to compare signal and background libraries of the Hfq and CsrA CLIP experiments in order to remove non-specific noise from the CLIP data and return the real protein binding sites (see Chapter 8).

1.8 Functional enrichments and pathway analyses

Often, biological experiments will return candidate lists that are difficult to evaluate as a whole, especially if the lists are long. Such experiments can be in vivo or in vitro based, or may come from predictive in silico algorithms. Common wet-lab approaches that return such lists are microarray or RNA-seq experiments for cells that have been subjected to different treatments or are of a different type[446]. In this context, for instance, a researcher may be comparing gene expression for WT cells and cells in which a specific gene has been knocked out or overexpressed. Upon knock out or overexpression it can be assumed that a subset of the genome is differentially expressed if the gene of interest is of any importance (see Section 1.7). In many cases this subset will contain an underlying pattern. However, without prior personal knowledge, it is initially challenging to tap into the information contained in the list. Here, functional enrichment and pathway analyses have proven to be valuable scientific assets. Among others the Gene Ontology (GO)[447] or Kyoto Encyclopedia of Genes and Genomes (KEGG)[448] databases are central hubs for the collection of gene functions and relationships. Such databases are the foundation for functional enrichment algorithms, which assess if there is a statistically significant over representation of specific functional terms in the list of interest as compared to a background (usually the whole genome). The over representation of terms may hint at what type of regulon a knocked out or overexpressed gene is serving. Computational approaches yielding ranked lists may be predictive algorithms such as the previously mentioned genome wide sRNA target predictions[144] or predictions for the binding sites of RBPs[282]. Here, functional enrichment analysis of the top predicted candidates can strongly aid in reducing the amount of false positive predictions. For sRNAs specifically, it has been shown that they often bind a set of mRNA targets that are connected via common functional terms such as amino acid metabolism, iron or carbohydrate transport (see Sections 1.4.2, 1.4.3, 1.4.4). If the algorithmic prediction succeeds in placing enough of the correct targets in the top list that is above a user defined score cutoff, subsequent functional enrichment of these top candidates can aid in reducing the overall number of interesting candidates (see Figure 19). Furthermore, similar functionally enriched terms retrieved from independent experiments on the same gene of interest can be a compelling source of scientific evidence[75].

Currently, at least 68 methods for this kind of analysis have been published. They can be subdivided into gene set enrichment analysis (GSEA), singular enrichment analysis (SEA) and modular enrichment analysis (MEA) methods[449]. While some of the implementations comprise only one of these classes, others belong to two.

GSEA was first introduced in 2003 by a study on human diabetes[450]. The major benefit of GSEA is that it does not need a predefined sublist of interest. This means arbitrarily set significance thresholds can be avoided and the entire list of genes that are considered in the experiment is used for enrichment analysis[449]. For this the list of genes needs to be sorted, for instance based on fold change of a signal versus control library. Then all gene sets or functional term lists are permuted and a running sum is calculated over all genes in the sorted experimental list for every term. A functional term itself is comprised of a list that contains entries of genes that are known to belong to this term. The running sum for an individual functional term is calculated by a scoring scheme that rewards an entry in the gene list if it belongs to the current functional term and penalizes an entry if it does not. In the end this gives rise to a running sum for every functional term that is being tested. The maximum of every running sum is then selected. This maximum value for a specific term is referred to as a term's enrichment score (ES). The maximum ES (MES) from all ES is then selected. To asses the statistical significance of the MES, a p-value can be calculated for it by shuffling the original gene list and calculating MES for the resulting randomized lists. This permutation and MES calculation has to be repeated an appropriate number of times ($\approx$1000) after which it gives rise to a plentiful distribution of background MES. The significance of the MES from the real data can then be assessed by calculating the fraction of MES in the background data that are larger than the real data MES, and the total number of background MES[450].

SEA methods, in contrast to GSEA approaches, work by first selecting a gene list of interest e.g., a set of differentially expressed genes in two experiments. For the candidates in this list a set of functional terms retrieved from a database can be permuted. While iterating over the individual terms, Fisher's exact test can be applied to calculate p-values and check if candidates of a specific term are statistically over represented in the gene list of interest when compared to the background gene list. Hence, a p-value is assigned to every functional term in the analysis and the functional terms with significant p-values can be reported and further investigated.

The statistical basis for MEA methods is the same as for SEA methods; however, they connect the terms used in SEA methods to larger more general groups, which can aid in a better understanding of the larger biological picture[454,455]. For example, a regulator may repress a set of enzymes that associate with a mixture of metals such as iron, zinc, copper and manganese. After an SEA analysis, all of these metals may show up as functionally enriched terms. While reporting these terms individually is not wrong, highlighting the universal term that joins these enzymes – metal binding – is more informative with respect to the regulator's overall function.

Clearly, the quality of every pathway or enrichment analysis is centrally dependent on the abundance of knowledge and annotation for an organism of interest, limiting such analyses to well understood species and pathways. This notwithstanding, the introduction of analysis platforms such as DAVID[451,452] and GSEA[456], which allow automated, holistic interpretation

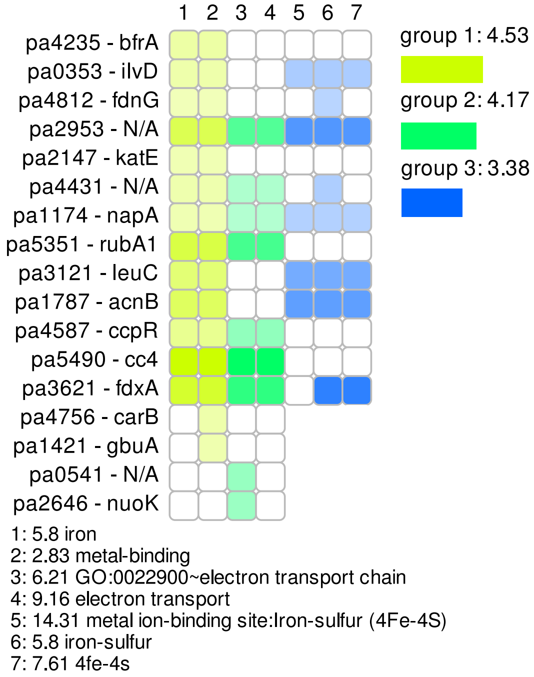


Figure 19. The figure shows a graphical representation of a functional enrichment analysis returned from DAVID-WS[451] using a CopraRNA[77,226] prediction list for the sRNA PrrF1[263] as input. Columns represent functional terms and rows represent genes. Different colors indicate different groups of functional terms. The DAVID enrichment score is indicated after the group number. A score of 1.3 or higher is considered to be statistically significant[452]. The identifiers on the left are locus tags and gene names from *Pseudomonas aeruginosa* PAO1 (NC_002516)[453]. Colored squares indicate genes that are connected to a functional term. White squares indicate the opposite. The color opacity of a specific square indicates the CopraRNA p-value. The functional terms related to the column numbers are shown at the bottom. The number before the term name indicates the terms' fold enrichment.

of experimental lists, represents a quantum leap in life science research.

2 Discussion

The most advanced state-of-the-art RNA-RNA interaction prediction algorithms, such as IntaRNA[176] and RNAup[413], available at the beginning of this thesis employ a thermodynamic energy model that also considers interaction site accessibility to predict hybrids between different RNA molecules. While this model is sound and readily able to predict the correct hybrids for pairs of RNA molecules known to interact, it often fails in predicting sRNA targets at the genomic scale due to a large number of false positive predictions clouding the pool of real targets. This serves to show that standard approaches in the field of RNA interaction prediction are not capable of sufficiently condensing the group of putative interaction partners for prokaryotic *trans*-acting sRNAs. As previously pointed out (see Section 1.6.6), predictive approaches can suffer from a lack of information on the entirety of crucial properties of in vivo systems, which can cause them to make predictions that at first glance infer an in vivo interaction that is actually not real.

Comparative prediction approaches with homologous sequences from distinct organisms are a common solution, but appear to be excessively restrictive and thus reduce the false positive rate at the cost of an inappropriately high increase of false negatives. This behavior may be attributed to the observation that sRNA target sites are not consistently conserved[457]. To this end, the development of the CopraRNA algorithm (see Chapter 4) aimed at reducing false positives without enforcing consensus rules too strict to model the majority of conceivable RNA-RNA interactions. Interestingly, the most successful strategy appears to be the most unrestrictive one possible. Specifically, CopraRNA only requires an interaction to be present anywhere on the homologous putative targets and does not score how conserved a particular interaction itself is. This means that the interactions predicted for the organisms participating in a CopraRNA analysis could be scattered over the full length of the putative target region, without being penalized for this. In practice, however, this extreme scattering can usually not be observed for confirmed interactions. A systematic comparison of CopraRNA with other leading sRNA target prediction tools[176,396,458] and microarray-based target prediction clearly showed that the newly developed algorithm represents the new in silico state-of-the-art with respect to prediction accuracy. This result was later confirmed by an independently conducted and comprehensive benchmark study on presently available sRNA target prediction algorithms[417]. In line with these results, CopraRNA should be a standard method employed at the beginning of every sRNA characterization study given that homologs of the sRNA of interest are available. In some cases it is apparent that CopraRNA coupled with functional enrichment analysis is capable of determining the correct regulon for certain sRNAs without any additional wet-lab-based data. One of many examples of this is depicted in Figure 19. Clearly, this is not always the case and even the most compelling predictions need to be followed up and confirmed with experimental data from the wet-lab.

It should be stated that the run times of CopraRNA predictions can be significantly higher than those of the single organism competitor algorithms. This is especially true if long sRNA sequences are used in target predictions including many organisms. In these cases CopraRNA predictions can take over 24 hours. While this is a downside of the algorithm, the compensation provided by the increased accuracy is significantly more important and the longer run time is thus justified.

The time period after the release of the CopraRNA algorithm showed its extended usefulness in internal projects that investigated sRNAs in species outside of the Enterobacteriaceae. This reinforces the results from the original CopraRNA study (see Chapter 4), which suggested that CopraRNA is not limited to Enterobacteriaceae alone. Thus, CopraRNA has been a valuable asset in the functional characterization of the sRNAs AbcR1 from *A. tumefaciens* (see Chapter 5), EcpR1 from *S. meliloti* (see Chapter 7), PsrR1[75] and NsiR4[76] both from *Synechocystis* sp. PCC 6803. Furthermore, several external projects have been successfully using CopraRNA[459–464]. These studies show that the algorithm has been actively adopted by the research community.

Because the improvement achieved by the original CopraRNA algorithm was so striking, it did not initially seem to call for any immediate improvements. This notwithstanding, the overall false positive rate of raw CopraRNA predictions (i.e. without automatic or manual post processing), as judged by currently confirmed sRNA-mRNA interactions, is still high. This led to the question, how including evidence from an additional factor – Hfq – important for RNA-RNA interaction mediation may be able to aid in further minimizing this issue. To this end, experimental transcriptome wide binding profiles for Hfq in *Salmonella* were determined and combined with CopraRNA target predictions. The results presented in Chapter 8 are indeed able to show that including experimental data from Hfq CLIP experiments can significantly reduce the false positive rate of CopraRNA, and hence lead to more informed selection of candidates for wet-lab-based confirmation of RNA-RNA interaction partners. Thus the putative interaction between the Spot42 sRNA and the *mglB* mRNA was selected as a promising, yet unknown candidate and subsequently confirmed experimentally. Even though this method allows for the selection of such high quality candidates for verification, it suffers from two pitfalls. Firstly, it is only viable for the investigation of RNA-RNA interactions that employ Hfq. Secondly, like microarray-based target prediction, it can only capture interactions between RNAs that are actively expressed in the applied experimental conditions. These two points show that adding organism and expression specific information to the predictive system can cause a loss of generality. While these pitfalls must be acknowledged, they do not invalidate the usefulness of extending target predictions with information from further sources. Rather, they show that such extensions to a generic prediction algorithm have to be individually tailored to fit the investigated situation.

Internal and external research projects conducted since the original release of CopraRNA have led to further, currently unpublished observations, extensions and results. Firstly, studies reported that several sRNA molecules of the same type can bind multiple sites within a single target mRNA[183,327]. These reports led to the proposition of including binding site multiplicity into the predictions, which did however not yield striking results and has since been abandoned.

Secondly, it has become apparent that including homologous sRNA sequences from more organisms (30 or more; the original benchmarks were conducted with a maximum of eight organisms) well distributed along the phylogenetic tree will return more robust predictions with more real targets and promising target candidates in the top lists.

Finally, certain interactions only appear to be conserved in a subset of organisms participating in the comparative analysis. An example is the FnrS-metE RNA pair, which judging by the interaction predictions is not conserved throughout the entire enterobacterial family[77,233]. This also means that standard comparative methods, including CopraRNA, will fail to report such interactions. Recent internal advances, in part, remedy this problem by removing single organism predictions with poor p-values before p-value combination. A second, currently unpublished approach to reduce this issue has been termed auxiliary enrichment analysis. This method compares the functional enrichment returned for the CopraRNA prediction with the functional terms in the top list of an individual whole genome IntaRNA prediction for the same sRNA. Candidates in the IntaRNA top list that fit functionally enriched terms in the CopraRNA prediction, but are not present in the CopraRNA top list are then reported as promising, non-conserved candidates. Both aforementioned approaches can lead to less conserved RNA-RNA interactions being elevated to higher ranks and push them into the spotlight of post-processing analyses. Next to the sole fact that relevant, less conserved interactions can be shifted into focus, this type of analysis also pinpoints putative hot spots for evolutionary diversity. Future studies should focus on the investigation of such putative hot spots. This may provide insights into the development of differences in sRNA regulatory networks in diverse species.

3 Outlook

At the beginning of this thesis many research projects in the sRNA field were still focusing on in depth characterization of individual sRNA regulators that had been discovered in RNA-seq driven studies on diverse organisms. While many such studies are still ongoing, the area is more recently transitioning away from these one by one investigations. A special focus are sRNAs that may rather code for short peptides instead of acting by directly pairing with target RNAs. Some of these candidates may also act as dual action sRNAs like the already characterized sRNA SgrS. Furthermore, scientists are turning to the investigation of RNA binding proteins that impact post-transcriptional networks[465,466]. Maybe most important are the current efforts to retrieve experimental RNA-RNA interactomes at the genomic scale in the wet-lab. Recent advances in this field have been made for both eukaryotes[467,468] and prokaryotes[208,469]. Once entirely harnessed and standardized, such methods should be able to supply comprehensive overviews of RNA-RNA interaction regulatory networks, thus relieving the necessity for investigating RNA regulators individually.

At first glance, the current development of wet-lab methods for the discovery of RNA-RNA interactomes poses the question of how useful and relevant genome wide target prediction will be in future. Clearly, the overall relevance of such approaches in the big picture will be reduced. However, just like for all other wet-lab methods, RNA-RNA interactomes will only be able to capture what is actively expressed. This means certain RNA-RNA interactions that are important in non-standard or not tested conditions will be lost. In line with this, experimental RNA-RNA interactomes can be used to discover the regulons of RNA-RNA interaction based regulators and genome wide target predictions can be employed alongside to extend the wet-lab results. Selection of promising candidates from the prediction lists will also be greatly facilitated if the general schemes for the regulated targets of specific sRNAs can be deduced from the interactome. Importantly, targets exclusively detected in the in silico prediction may be able to aid in the design of additional interesting physiological conditions for the retrieval of experimental interactomes. Furthermore, the previously outlined prediction and investigation of evolutionary differences between RNA-RNA interactions in different species will most likely remain an in silico domain for the time being.

4 Comparative genomics boosts target prediction for bacterial small RNAs

Patrick R. Wright, Andreas S. Richter, Kai Papenfort, Martin Mann, Jörg Vogel, Wolfgang R. Hess, Rolf Backofen and Jens Georg (2013) **Proceedings of the National Academy of Sciences**, 110, E3487-E3496.
DOI: 10.1073/pnas.1303248110

Personal contribution

I implemented the software (CopraRNA) for this project, and helped in setting up the webserver. Furthermore, I contributed to the analysis and interpretation of the results and was involved in writing the manuscript.

Patrick R. Wright

The following co-authors confirm the above stated contribution.

Dr. Jens Georg

Prof. Dr. Rolf Backofen

Comparative genomics boosts target prediction for bacterial small RNAs

Patrick R. Wright[a,b], Andreas S. Richter[b], Kai Papenfort[c,d], Martin Mann[b], Jörg Vogel[c], Wolfgang R. Hess[a,e], Rolf Backofen[b,e,f,g,1], and Jens Georg[a,1]

[a]Genetics and Experimental Bioinformatics, Faculty of Biology, [e]Centre for Biological Systems Analysis, and [f]BIOSS Centre for Biological Signalling Studies, University of Freiburg, D-79104 Freiburg, Germany; [b]Bioinformatics Group, Department of Computer Science, University of Freiburg, D-79110 Freiburg, Germany; [c]Institute for Molecular Infection Biology, University of Würzburg, D-97080 Würzburg, Germany; [d]Department of Molecular Biology, Princeton University, Princeton, NJ 08544; and [g]Center for Non-Coding RNA in Technology and Health, University of Copenhagen, DK-1870 Frederiksberg C, Denmark

Edited by Gisela Storz, National Institutes of Health, Bethesda, MD, and approved July 30, 2013 (received for review February 22, 2013)

Small RNAs (sRNAs) constitute a large and heterogeneous class of bacterial gene expression regulators. Much like eukaryotic microRNAs, these sRNAs typically target multiple mRNAs through short seed pairing, thereby acting as global posttranscriptional regulators. In some bacteria, evidence for hundreds to possibly more than 1,000 different sRNAs has been obtained by transcriptome sequencing. However, the experimental identification of possible targets and, therefore, their confirmation as functional regulators of gene expression has remained laborious. Here, we present a strategy that integrates phylogenetic information to predict sRNA targets at the genomic scale and reconstructs regulatory networks upon functional enrichment and network analysis (CopraRNA, for Comparative Prediction Algorithm for sRNA Targets). Furthermore, CopraRNA precisely predicts the sRNA domains for target recognition and interaction. When applied to several model sRNAs, CopraRNA revealed additional targets and functions for the sRNAs CyaR, FnrS, RybB, RyhB, SgrS, and Spot42. Moreover, the mRNAs *gdhA, lrp, marA, nagZ, ptsI, sdhA*, and *yobF-cspC* were suggested as regulatory hubs targeted by up to seven different sRNAs. The verification of many previously undetected targets by CopraRNA, even for extensively investigated sRNAs, demonstrates its advantages and shows that CopraRNA-based analyses can compete with experimental target prediction approaches. A Web interface allows high-confidence target prediction and efficient classification of bacterial sRNAs.

regulatory RNA | *E. coli* | RNA–RNA interaction

Small RNAs (sRNAs) are ubiquitous and important regulators of gene expression in bacteria. The most common and best investigated *trans*-acting sRNAs regulate their targets posttranscriptionally by RNA–RNA interactions, often depending on the RNA chaperone Hfq (1). Individual functions of model sRNAs have been discovered primarily through extensive experimental work and may be assigned to many different stress responses and signal transduction pathways, covering virtually all aspects of bacterial growth (1, 2) and virulence (3). One of the most intriguing conceptual advances has been the identification of sRNAs as posttranscriptional regulators that act globally within complex regulatory networks. Examples for such sRNAs are GcvB, which is a major regulator of amino acid metabolism and directly controls ~1% of all *Salmonella enterica* mRNAs (4); MicA and RybB, which together constitute the repressor arm of the Sigma E response (5); and Spot42, a global regulator of catabolite repression (6). With the advent of high-throughput sequencing and comprehensive transcriptome analysis techniques, increasing numbers of new sRNAs have been detected in bacteria belonging to diverse taxa (7, 8). However, the experimental testing and verification of sRNA targets is costly, labor intensive, and may be challenging, even in model organisms. Moreover, for most environmentally and biotechnologically relevant microbes, experimental verification is hindered further by the lack of systems for their genetic manipulation.

The reliable computational prediction of sRNA targets promises a great reduction of required wet-laboratory analyses while enabling large-scale sRNA–mRNA network analyses in genetically intractable species. However, reliable in silico prediction of mRNA targets has been challenging because of the extreme heterogeneity of sRNAs in size, structure, and the typically short and imperfect sRNA–target complementarity (9). The existing tools for the genome-scale prediction of sRNA targets evaluate the strength of a particular sRNA–target interaction by either base pair complementarity (10) or thermodynamic models (11–13). The latter are built on the observed exponential correlation between repression strength and hybridization free energy (14), which can be corrected by an energy term that reflects the accessibility of the interaction sites (11, 12). However, despite continuous improvement of target prediction methods (15), even the most accurate methods integrating interaction site accessibility scoring and additional features, such as seed regions, produce many false positives and, thus, compromise the selection of putative targets for subsequent experimental investigation (16, 17).

Furthermore, the implementation of seed sequence conservation to improve sRNA target prediction has been difficult to achieve for bacterial systems because of the great flexibility of the interaction patterns (16). It is conceivable that the interaction is preserved while the actual interaction site is not. Therefore, to predict conserved interactions, it is necessary to combine evidence for interactions in different species without resorting to a consensus interaction-based approach.

Here, we introduce a computational approach that uses phylogenetic information from an extended model of sRNA–target evolution (CopraRNA, for Comparative Prediction Algorithm

Significance

This study presents a unique approach (CopraRNA, for Comparative Prediction Algorithm for sRNA Targets) towards reliably predicting the targets of bacterial small regulatory RNAs (sRNAs). These molecules are important regulators of gene expression. Their detailed analysis thus far has been hampered by the lack of reliable algorithms to predict their mRNA targets. CopraRNA integrates phylogenetic information to predict sRNA targets at the genomic scale, reconstructs regulatory networks upon functional enrichment and network analysis, and predicts the sRNA domains for target recognition and interaction. Our results demonstrate that CopraRNA substantially improves the bioinformatic prediction of target genes and opens the field for the application to nonmodel bacteria.

Author contributions: A.S.R., R.B., and J.G. designed research; P.R.W., K.P., and J.G. performed research; P.R.W., A.S.R., M.M., and J.G. contributed new reagents/analytic tools; P.R.W., W.R.H., and J.G. analyzed data; and P.R.W., A.S.R., K.P., J.V., W.R.H., R.B., and J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: jens.georg@biologie.uni-freiburg.de or backofen@informatik.uni-freiburg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1303248110/-/DCSupplemental.

for sRNA Targets). CopraRNA depends solely on the conservation of target genes (i.e., conservation of target regulation) and does not require conservation of specific interaction sequences (*SI Appendix*, Figs. S1 and S2).

By introducing a generic approach combining predictions for homologous targets in distinct organisms, we reduced the hitherto existing high false positive rate (FPR) of single-organism target prediction. Using this strategy, CopraRNA matches microarray-based experimental sRNA target prediction with respect to the number of correctly identified direct targets (Fig. 1*B* and Table 1) and the characterization of physiological functions of these sRNAs. Thus, it constitutes a significant improvement of in silico sRNA target prediction and enables competitive and functional large-scale initial screening for sRNA targets without experimental effort and costs. Application of CopraRNA to previously characterized sRNAs proposed and partially verified additional targets and functions for the sRNAs cyclic AMP activated sRNA (CyaR), FNR regulated sRNA (FnrS), RybB, RyhB, sugar transport-related sRNA (SgrS), and Spot42. Also, it suggested the *gdhA, lrp, marA, nagZ, ptsI, sdhA*, and *yobF-cspC* mRNAs as hubs targeted by up to seven different sRNAs. A Web interface for CopraRNA has been set up under http://rna.informatik.uni-freiburg.de/CopraRNA/.

Results

Prediction Strategy. CopraRNA begins with a genome-wide target prediction (12) for each considered organism, as summarized in Fig. 1*A*. The interaction energies are fitted to a general extreme value distribution and transformed into *P* values to normalize for organism-specific GC-content and dinucleotide frequency. These *P* values are combined for orthologous genes into a single *P* value per conserved interaction. Orthologous genes are determined based on the respective amino acid sequences (25); genes that are present in less than 50% of the investigated genomes are discarded. Two aspects require specific normalization. First, CopraRNA normalizes for the degree of overall dependency to account for the nonindependent *P* values that result from the

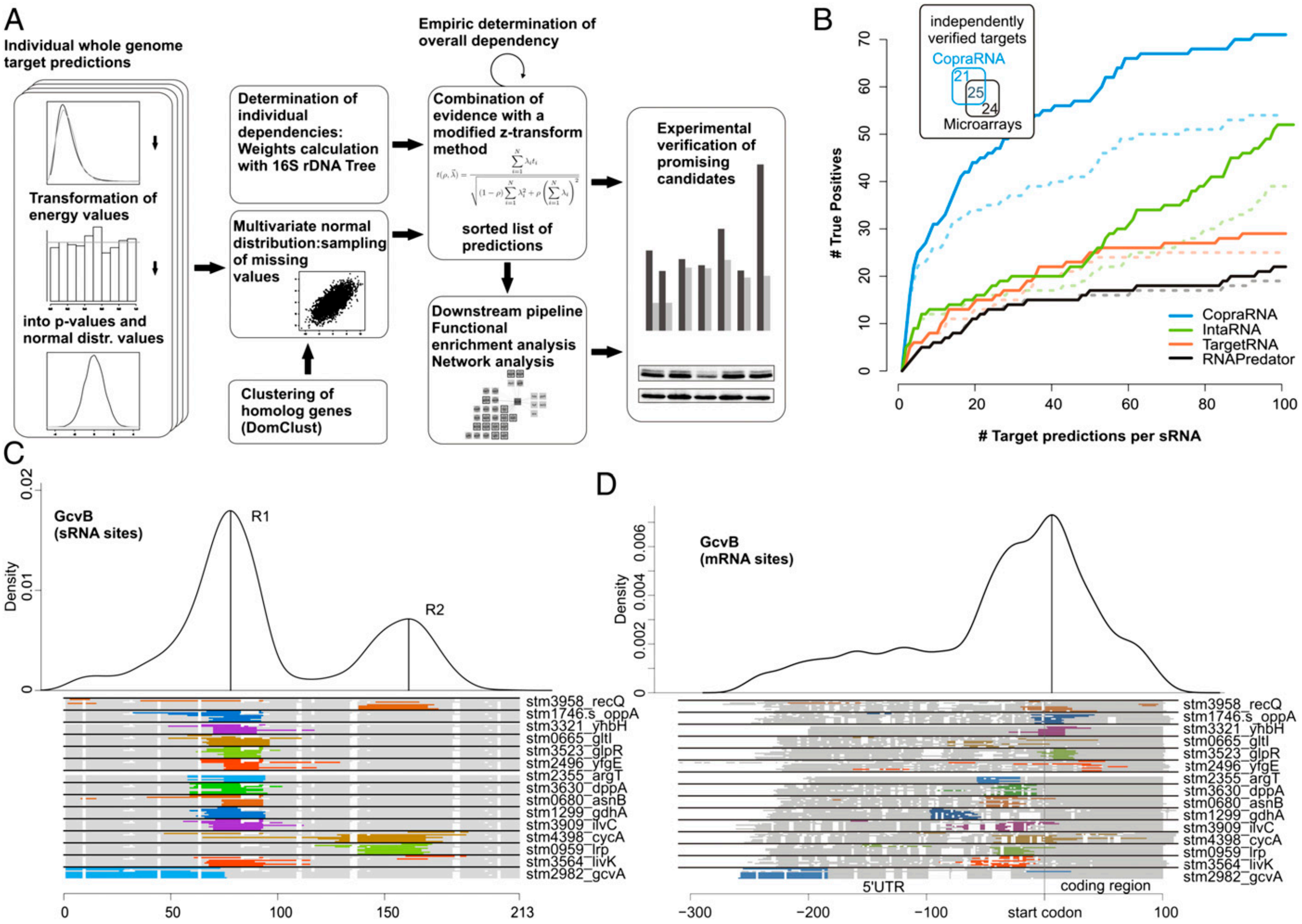


Fig. 1. (*A*) Schematic overview of the CopraRNA pipeline. (*B*) Comparison of CopraRNA predictions with microarray results and other target prediction methods. Genome-wide target predictions for 18 sRNAs in *E. coli* and *S. enterica* with 101 experimentally verified targets from the literature. The plot shows the number of correctly predicted targets (true positive predictions, *y* axis) vs. the number of target predictions per sRNA (*x* axis) for our comparative method CopraRNA and the existing single-organism–based methods IntaRNA, TargetRNA, and RNApredator. The results, including the verifications from this study, are shown with solid lines, and the results based on the benchmark set only are demarcated with a dashed line. (*Inset*) total numbers of independently verified targets detected by either CopraRNA (46 targets) or microarray experiments (49 targets) for the sRNAs CyaR, FnrS, GcvB, MicF, RyhB, SgrS, and Spot42; 25 targets were identified by both methods. The numbers refer to our benchmark dataset (*SI Appendix*, Table S1) and to the table comparing CopraRNA with different microarray experiments (Table 1). Visualization of the predicted interaction domains in GcvB (*C*) and the predicted mRNA targets of GcvB (*D*). The density plots at the top give the relative frequency of a specific sRNA or mRNA nucleotide position in the predicted sRNA–target interactions. The plots combine all predictions with a *P* value ≤0.01 in all included homologs. Local maxima indicate distinct interaction domains and are marked with upright lines. The schematic alignment of homologous sRNAs and targets at the bottom show the predicted interaction domains. The aligned regions are displayed in gray, gaps in white, and predicted interaction regions in color (color differences are for contrast only). The locus tag and gene name (if available) of a representative cluster member are given on the right.

Table 1. Comparison of CopraRNA predictions and published microarray studies

	CopraRNA			Microarray				
sRNA	No. of candidates ($P \leq 0.01$)	No. of candidates after postprocessing*	No. verified[†]	No. sig. diff. expr. genes[‡]	No. verified[†]	Ref.	No. overlap[§] verified/ unverified	Overlap genes[¶]
CyaR	69	55	1 + 3[‖]	24 genes	4	18	1/1	*fepA*, ***ompX***
				1 gene	1	19	1/0	***ompX***
FnrS	67	41	3 + 4[‖]	16 genes/11 operons	6 + 1[‖]	20	3/0	***marA***, ***sodB***, ***yobA***
				31 genes	7 + 1[‖]	21	4/2	*adhP*, ***marA***, ***sfcA/maeA***, ***sodB***, *ydhD/grxD*, ***yobA***
GcvB	60	34	14	54 genes	16	4	10/3	***argT***, *aroP*, ***brnQ***, ***cycA***, ***dppA***, ***gdhA***, ***gltI***, ***lrp***, ***oppA***, ***serA***, ***sstT***, *trpE*, *yifK*
MicF	50	30	4	5 genes	4	22	2/0	***lrp***, ***ompF***
RyhB	70	37	2 + 5[‖]	56 genes/18 operons	3 + 1[‖]	23	3/3	*frdA*, ***fumA***, *msrB*, ***sdhA***, *sdhD*, ***sodB***
SgrS	66	35	2 + 1[‖]	6 genes	4	24	2/0	***ptsG***, ***yigL***
Spot42	85	48	4 + 3[‖]	16 genes	7	6	3/0	***galk***, ***gltA***, ***xylF***

The candidates after postprocessing for these sRNAs are given in Table S5.

*Top 15 targets + automatically and manually functionally enriched.

[†] Verified targets after postprocessing regarding the benchmark list (*SI Appendix*, Table S1), published data, and this study.

[‡] Significantly differentially expressed genes with regard to the respective publications.

[§] Genes detected by prediction and microarray (independently verified/unverified).

[¶] Independently verified targets are in boldface.

[‖] Verified in this study.

general sequence conservation between related organisms. Second, the individual dependencies have to be calculated because, in most cases, the considered organisms will not be equidistant from each other. Thus, we additionally used species-specific weights that were calculated based on 16S rDNA-based phylogenetic trees. The combination of the *P* values used a modified z-transform method, which permits adjustment for dependency in the data and a weighting based on the phylogenetic relationship (26). We defined significance thresholds either on CopraRNA *P* values or on q-values (27); the latter provide correction for multiple testing by controlling the false discovery rate (FDR). Both methods have proven useful for the analysis of the benchmark dataset. The chosen *P* value threshold of 0.01 allows for the detection of approximately half of all verified benchmark targets (*SI Appendix*, Fig. S3*A*) and was applied for the functional enrichment and network analysis. The q-value gives a measure of how many false positive predictions are expected in the group of targets called significant. True positives are all experimentally verified targets (with regard to our benchmark dataset in *SI Appendix*, Table S1) within the positive predictions, whereas false positives are all positive predictions that are no real targets, i.e., in our case, those that have not been verified experimentally. Positive predictions (also called candidates below) are all targets that match the respective threshold criterion (e.g., a *P* value ≤0.01 or a given rank); they consist of true positive and false positive predictions (statistical terms are defined also in *SI Appendix*). A reliable bioinformatic prediction tool for sRNA targets should not predict more than ~50% of false positive targets; therefore, we chose a q-value threshold of 0.5. The validity of this approach for CopraRNA was tested with the prediction for GcvB. We assume that GcvB, with its 22 verified targets, is so far the most thoroughly investigated sRNA (4). In the CopraRNA prediction of GcvB, 37 targets are predicted with a q-value ≤0.5. Of these, 35 have homologs in *Escherichia coli* or *S. enterica*, 11 of which have been verified. Fifteen of the 35 homologs are involved in amino acid metabolism or transport, i.e., they fit to the known biological function of GcvB. This corresponds to an FDR of 69% or 57%, respectively, with regard to currently known targets and is not very far from the statistical estimate of 50%. In general, the number of significant predictions with a q-value ≤0.5 is a rough approximation of the expected number of targets and the prediction quality of the tested sRNA. A detailed description of the CopraRNA procedure is provided in *SI Appendix*.

Benchmark with Experimentally Verified Targets. To evaluate the accuracy of CopraRNA, we performed a benchmarking test on a set of 18 conserved enterobacterial sRNAs and their 101 experimentally verified mRNA targets (modified from ref. 16) using homologous sequences from three to eight organisms (*SI Appendix*, Fig. S4). Compared with predictions by the existing approaches IntaRNA (12), TargetRNA (10), and RNApredator (11) (Fig. 1*B*), CopraRNA showed a clear improvement in the sensitivity or true positive rate (sensitivity $= \frac{\#\text{ true positives}}{\#\text{ positives}}$) and positive predictive value (PPV $= \frac{\#\text{ true positives}}{\#\text{ positive predictions}}$). Based on published data, CopraRNA's top 1 target predictions were correct for 8 of 18 sRNAs (PPV: 44%), compared with 5 (PPV: 28%) for IntaRNA, 2 (PPV: 11%) for TargetRNA, and 1 (PPV: 6%) for RNApredator. When considering the top 5 and top 15 target predictions per sRNA, CopraRNA correctly detected 23 and 32, respectively, of all 101 targets (true positive rate: 23% and 32%, respectively), which constitutes a twofold increase in sensitivity compared with IntaRNA and a 2.9-fold and fourfold improvement compared with TargetRNA and RNApredator, respectively (*SI Appendix*, Table S2). In addition, our experimental verification (below) demonstrated that the existing lists of known targets are still incomplete, implying an underestimation of the true positive rate (Fig. 1*B*).

In many cases, the comparative approach resolved the problem of false negatives (i.e., verified targets missed in the prediction) in single-organism–based methods. Prominent examples are the GcvB targets *lrp* (4), *oppA* (4), and *stm3903* (4); the RybB target *ompN* (28); and the Spot42 target *gltA* (6). The ranking of these targets improved from rank 95 to 3, rank 164 to 14, rank 1,297 to 40, rank 69 to 3, and rank 392 to 2, respectively (*E. coli*- or *S. enterica*-specific prediction vs. CopraRNA prediction). The benchmark dataset and the complete ranked list of all predictions are given in *SI Appendix*, Table S1 and Table S3.

Prediction of Interaction Domains. In addition to the ranked list of predicted targets, CopraRNA provides comparative information on the putative interaction sites of the sRNA and its mRNA targets. These data are summarized in two density plots combining all predictions with a *P* value ≤0.01 for a specific sRNA (Fig. 1 *C* and *D* shows the GcvB example). Based on multiple sequence alignments, these plots visualize the frequency of single residues participating in the predicted sRNA–mRNA interactions. The plots are complemented by a series of schematic alignments for both sRNAs and mRNAs that highlight organism-specific predicted interactions. From these plots, the interaction domains of the sRNA can be inferred, as they provide the combined information of accessibility, complementarity, and phylogenetic conservation.

This visualization immediately highlights the two previously described interaction regions of GcvB (4) (Fig. 1*C*), the three different interaction regions of Spot42 (6), and the single 5′ located region of RybB (9) (*SI Appendix*, Fig. S5). In agreement with the published data for Spot42, *gltA* is targeted by the first single-stranded region (6) centered at position 6 in the multiple-sequence alignment (*SI Appendix*, Figs. S5 and S6). The newly identified targets *sucC* and *gdhA* base pair with the second and third interaction region of Spot42, respectively. For *galK*, all three regions are predicted to be involved in the interaction for four of the eight investigated organisms (*SI Appendix*, Fig. S5). As previously described (4), GcvB targets *lrp* and *cycA* via region "R2" of the sRNA (Fig. 1*C*), whereas most targets (e.g., *dppA* and *oppA*) interact with region "R1." In the case of RprA, the full-length form appears to have two interaction domains, and only the distal site is retained after processing (29) (*SI Appendix*, Fig. S5), leading to a significant shift in the list of predicted targets. The mRNA plots are useful to obtain a rapid overview on the predicted interaction sites regarding their relative position and their phylogenetic conservation. The density plot also reveals the predominant interaction regions when using target sequences of the same length. For GcvB targets, there is a clear tendency toward the region near the start codon (Fig. 1*D*).

Functional Enrichment of Predicted Targets. Many well-studied sRNAs control sets of functionally related genes [e.g., RyhB, nonessential iron-binding proteins (30), GcvB, amino acid biosynthesis genes (4)]. Therefore, we analyzed the top-ranked targets of all benchmark sRNAs for functional relationships based on automated functional enrichment using the database for annotation, visualization, and integrated discovery (DAVID) (31). A combination of CopraRNA and functional enrichment provided very clear results for several sRNAs and suggested their potential involvement in diverse cellular networks (Tables S4 and S5). The DAVID Web server clusters related terms and calculates a combined enrichment score. Table 2 shows representative terms for the most strongly enriched clusters of selected sRNAs. The accuracy of this approach is demonstrated exemplarily for GcvB: this sRNA has a broad set of 22 verified target mRNAs (4) and a clearly defined function as a regulator of amino acid metabolism and transport (4). GcvB has 60 positive predictions (*P* value ≤0.01, *E. coli*). Seven experimentally verified targets are in the top 10 list, which supports the prediction accuracy of our algorithm and represents a PPV of 70%. Among the 60 candidate targets, 19 were annotated with the term "cellular amino acid biosynthetic process" and were significantly enriched (enrichment score ∼6.65) over background (i.e., all genes included in the prediction output). In summary, 26 of the 60 predictions were grouped as amino acid related, including genes for 11 amino acid biosynthesis proteins, 9 amino acid transporters, and 4 peptide transporters. These results are complementary

Table 2. Results of the functional enrichment analysis using the DAVID Web server (31)

sRNA	No. predicted	Enrichment score	Category	Term	No.
CyaR	69	4.95	UP_SEQ_FEATURE	Topological domain:Periplasmic	26
		3.45	SP_PIR_KEYWORDS	Cell inner membrane	32
		2.15	GOTERM_BP_FAT	GO:0005976∼polysaccharide metabolic process	11
FnrS	67	2.43	SP_PIR_KEYWORDS	Flavoprotein	6
		1.44	GOTERM_MF_FAT	GO:0005506∼iron ion binding	9
		1.41	GOTERM_MF_FAT	GO:0046872∼metal ion binding	19
GcvB	60	6.65	GOTERM_BP_FAT	GO:0008652∼cellular amino acid biosynthetic process	19
		4.12	GOTERM_BP_FAT	GO:0006865∼amino acid transport	9
		2.78	GOTERM_MF_FAT	GO:0015171∼amino acid transmembrane transporter activity	5
MicA	46	1.97	GOTERM_CC_FAT	GO:0009279∼cell outer membrane	6
		1.12	GOTERM_BP_FAT	GO:0000271∼polysaccharide biosynthetic process	6
MicF	50	2.36	GOTERM_CC_FAT	GO:0044462∼external encapsulating structure part	7
		2.14	GOTERM_CC_FAT	GO:0030312∼external encapsulating structure	16
		1.28	SP_PIR_KEYWORDS	Lipoprotein	5
RyhB	70	3.41	GOTERM_MF_FAT	GO:0005506∼iron ion binding	13
		2.86	GOTERM_MF_FAT	GO:0046872∼metal ion binding	22
		2.59	GOTERM_MF_FAT	GO:0051536∼iron-sulfur cluster binding	9
SgrS	66	1.62	KEGG_PATHWAY	02060: phosphotransferase system (PTS)	5
		1.36	GOTERM_MF_FAT	GO:0046872∼metal ion binding	17
		1.35	GOTERM_BP_FAT	GO:0051188∼cofactor biosynthetic process	7
Spot42	85	2.96	GOTERM_BP_FAT	GO:0046356∼acetyl-CoA catabolic process	7
		2.53	GOTERM_BP_FAT	GO:0006732∼coenzyme metabolic process	12
		1.83	KEGG_PATHWAY	00020:Citrate cycle, tricarboxylic acid cycle (TCA cycle)	5
FsrA	54	4.77	GOTERM_MF_FAT	GO:0051536∼iron-sulfur cluster binding	8
		3.81	GOTERM_BP_FAT	GO:0022900∼electron transport chain	6
		3.69	UP_SEQ_FEATURE	domain:4Fe-4S ferredoxin-type 2	4
PrrF	103	4.47	GOTERM_MF_FAT	GO:0051536∼iron-sulfur cluster binding	12
		4.88	GOTERM_MF_FAT	GO:0005506∼iron ion binding	20
		3.81	SP_PIR_KEYWORDS	electron transport	7
SR1	50	1.88	GOTERM_BP_FAT	GO:0030435∼sporulation resulting in formation of a cellular spore	8

The top 3 significantly enriched terms (DAVID enrichment score ≥1.1) for 11 tested sRNAs are shown. For each sRNA, the number of predicted targets with a *P* value ≤0.01 (column 2), the score of the enriched functional cluster (column 3), the name and source of a representative term of this cluster (columns 4 and 5), and the number of unique genes in this cluster (column 6) are given. Individual gene members of the enriched terms are given in Table S5.

to the existing experimental findings and add several plausible candidates.

The known functions of GcvB were predicted almost completely by CopraRNA and the subsequent functional enrichment. The top 15 predictions and functionally enriched target candidates are shown in Fig. 2*A*.

CopraRNA also returned the correct functional characterization for several other sRNAs. The predicted targets of MicA (Table 2 and *SI Appendix*, Fig. S7) and MicF were strongly enriched for outer membrane proteins, whereas the most strongly enriched cluster of RyhB targets consists of iron-binding proteins (Table 2 and Fig. 2 *B* and *C*).

Network Analysis of Predicted Targets. Certain genes serve as regulatory hubs and are targeted by several sRNAs. For example, the mRNA encoding the alternative sigma factor RpoS is targeted directly by at least three sRNAs, the Arc-associated sRNA Z (ArcZ), DsrA, and the RpoS regulator RNA (RprA) (1), whereas the *csgD* mRNA is regulated by five different sRNAs, i.e., GcvB (32), the multicellular adhesive sRNA (McaS) (32, 33), the OmpR-regulated sRNA A/B (OmrA/B) (34), and RprA (35). Computational target prediction by CopraRNA allows the analysis of a high number of sRNAs, and the results can be combined to infer the gene regulatory network for a given organism. Indeed, our global network analysis based on the benchmark dataset predicted known and potential hotspots of sRNA-based regulation. In total, 15 mRNAs were predicted to be targeted by four or more sRNAs and ~50 mRNAs by three or more sRNAs (Table S6). A striking example of an mRNA with multiple potential sRNA regulators encodes Lrp (leucine-responsive regulatory protein) and is predicted to be regulated by 7 of the 18 investigated sRNAs, including the previously identified regulators MicF (22, 36) and GcvB (4). The mRNA encoding the succinate dehydrogenase subunit SdhA has six predicted sRNA

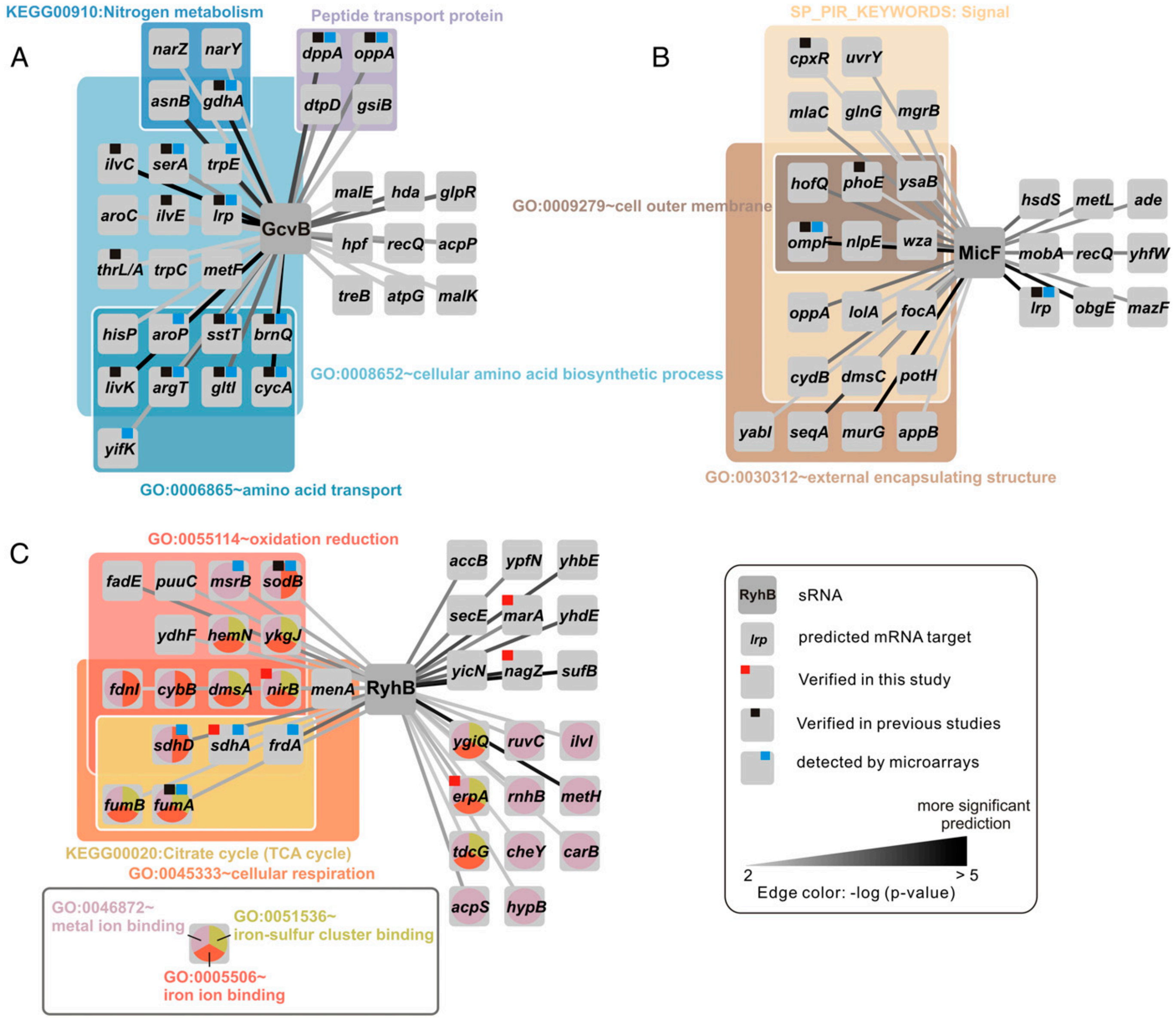


Fig. 2. Visualization of the functional enrichment analysis. All top 15 target predictions are shown plus predictions with a CopraRNA *P* value ≤0.01 that are functionally enriched (selected enriched terms). The edges connecting the sRNAs and targets are color coded according to the CopraRNA prediction *P* value, a darker color indicates a statistically more significant prediction. Previously experimentally verified targets from the literature [with regard to our benchmark list (*SI Appendix*, Table S1)] are marked with a black square, verifications from this study with a red square, and targets detected by microarrays with a blue square. Functionally enriched targets are color coded with respect to the enriched term. Results for (*A*) GcvB, (*B*) MicF, and (*C*) RyhB.

regulators, three of which were verified in this study (see below). We also detected multiple regulators of *csgD* and *rpoS* mRNAs. In addition to OmrA/B (34) and RprA (35), we predicted ChiX as a potential regulator of *csgD*. Another interesting example is the *yobF-cspC* dicistron with four potential regulators (CyaR, OmrA/B, and OxyS). From these, OxyS was previously shown to negatively regulate the *yobF-cspC* mRNA (10). The network obtained for 18 sRNAs and their previously verified and new targets is presented in Fig. 3*A*. In total, when using a *P* value threshold of 0.01, CopraRNA predicted 52 of the 101 benchmark targets. Furthermore, we verified 17 as yet unknown targets, uncovering connections between the regulatory networks of GcvB and Spot42, CyaR, RyhB and FnrS, and CyaR and SgrS. FnrS and RyhB share a dense overlapping regulon of at least four targets (Fig. 3*A*). Additionally, several operons were predicted to be influenced by multiple sRNAs: the *sdhCDAB-sucABCD* operon is targeted by five sRNAs at three different positions (Fig. 3*B*); Spot42 and RyhB each regulate two genes in the operon, *sdhC* (37) and *sucC*, as well as *sdhD* (37) and *sdhA*, respectively. In addition, the *iscRUAB* operon is regulated by both FnrS and RyhB (38) (Fig. 3*C*).

Experimental Verification of Predicted Targets. Based on the benchmark results, we restricted the final set of target candidates for each sRNA to the top 15 predictions plus candidates that belong to the functional-enriched terms (Table S5). This approach provides a reasonable balance between sensitivity and specificity because it uses the high positive predictive value in the topmost predictions (*SI Appendix*, Fig. S3*B*) while allowing investigation of an extended target set. We selected 23 previously uncharacterized potential targets (*SI Appendix*, Table S7) for experimental testing using a GFP reporter system tailored to investigate posttranscriptional regulation (22). We verified 17 additional targets, which equals a success rate of ∼74%, and exemplarly proved the predicted interaction sites of *yobF*-CyaR, *iscR*-FnrS, *nirB*-RyhB, and *gdhA*-Spot42 through the introduction of compensatory mutations and for *marA*-FnrS, *erpA*-RyhB, *marA*-RyhB, and *sucC*-Spot42 by point mutations in their respective 5′UTRs (Fig. 4 *A* and *B* and *SI Appendix*, Fig. S8). Interestingly, the point mutations in the *marA*^{*1} construct resulted in an increased repression by wild-type RyhB, which indicates an improved RNA–RNA hybrid formation. Posttranscriptional repression of the remaining predicted targets was tested by flow cytometry (Fig. 4*C*) or Western blots (*SI Appendix*, Fig. S9). An overview of the constructs used and the respective mean fluorescence intensities is given in *SI Appendix*, Figs. S9 and S10. Most of the predicted interactions resemble the classic binding proximal to the translational start site. However, the binding sites for Spot42 in *gdhA* and *icd* align with positions +80 and +75 downstream from the start codon, deeply within the

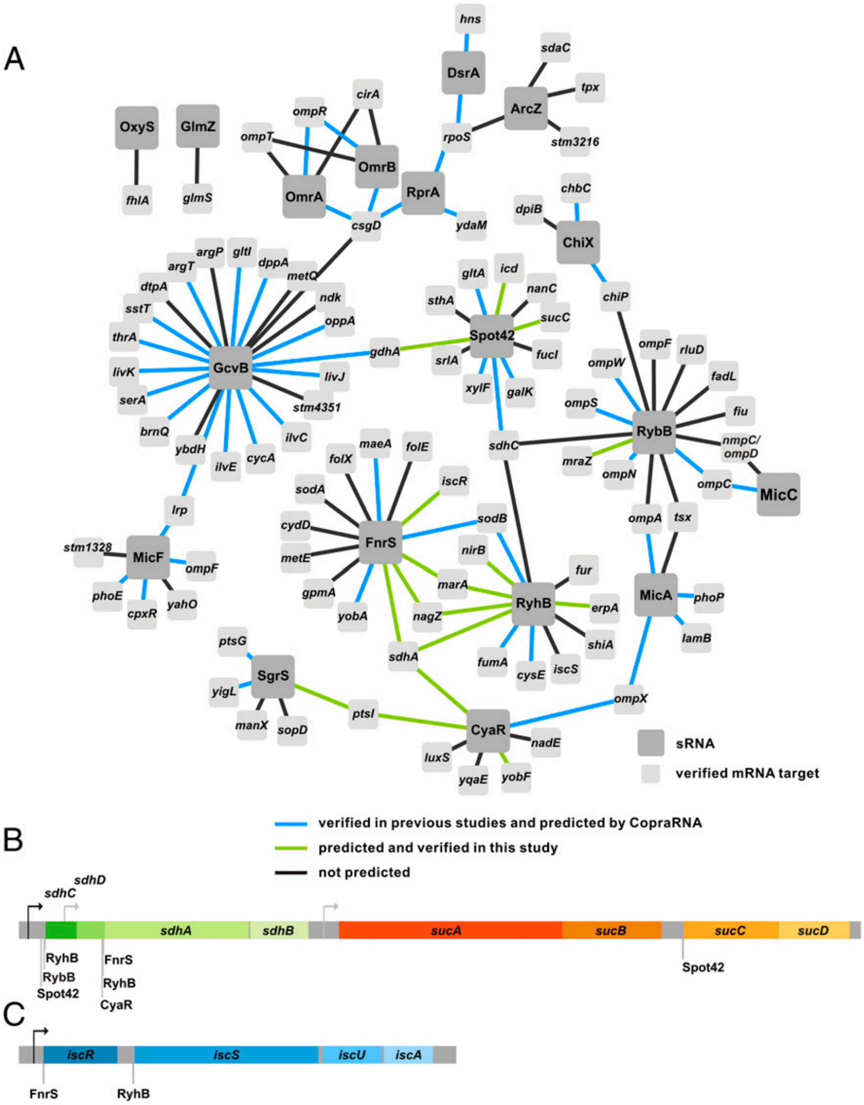


Fig. 3. (*A*) Network of verified targets for the 18 sRNAs of the benchmark dataset. Visualization of the (*B*) *sdhCDABsucABCD* and (*C*) *iscRSUAB* operon with verified interaction sites; the promoters are annotated according to EcoCyc (52).

coding region. A direct inhibition of translation seems unlikely for these targets; rather, we assume a mechanism that reduces the half-life of the mRNAs, as shown for the *ompD*–MicC interaction in *S. enterica* (39, 40).

Performance of CopraRNA for sRNAs from Nonenterobacterial Species. To evaluate the performance of CopraRNA for sRNAs that are not conserved in *E. coli* or *S. enterica*, we extended our benchmark dataset by five additional sRNAs from a wide range of bacterial families and phyla—the Fur-regulated sRNA A (FsrA) and SR1 (Firmicutes, Bacillaceae), LhrA (Firmicutes, Listeriaceae), the inhibitor of hctA translation (IhtA) (Chlamydiae), and PrrF (Proteobacteria, Pseudomonadaceae)—with a total of 17 experimentally verified targets (*SI Appendix*, Table S8). CopraRNA detects 11 of the 17 verified targets in the top 35 predictions, which resembles a true positive rate of ∼65% and a PPV of ∼6.3%. Again, this is at least ∼3.7 times better than the single-organism–specific methods (*SI Appendix*, Fig. S11). We also obtained intriguing functional enrichments for FsrA and PrrF (Table 2 and Table S5). The topmost enriched term for the predicted FsrA and PrrF targets is "GO:0051536∼iron-sulfur cluster binding" followed by other iron-related terms. This is in agreement with the known roles of these sRNAs in the iron stress response (30) and may hint at additional yet-unknown target genes of those sRNAs. The complete prediction dataset is given in Table S9.

Discussion

Comparison with Other Target Identification Strategies. In this study, we present a comparative method for sRNA target identification in bacteria. The method is superior to existing bioinformatics tools (Fig. 1*B*) and works for a wide range of bacterial organisms. For seven tested benchmark sRNAs, CopraRNA can compete with microarray-based experiments for target detection (Table 1). CopraRNA is available as an easy-to-use Web interface (http://rna.informatik.uni-freiburg.de/CopraRNA/). True positive predictions are enriched by the downstream refinement of the prediction results through integration of existing data.

Using CopraRNA, we detected 17 as yet unknown targets for six sRNAs (Fig. 4 and *SI Appendix*, Fig S9). For the sRNAs

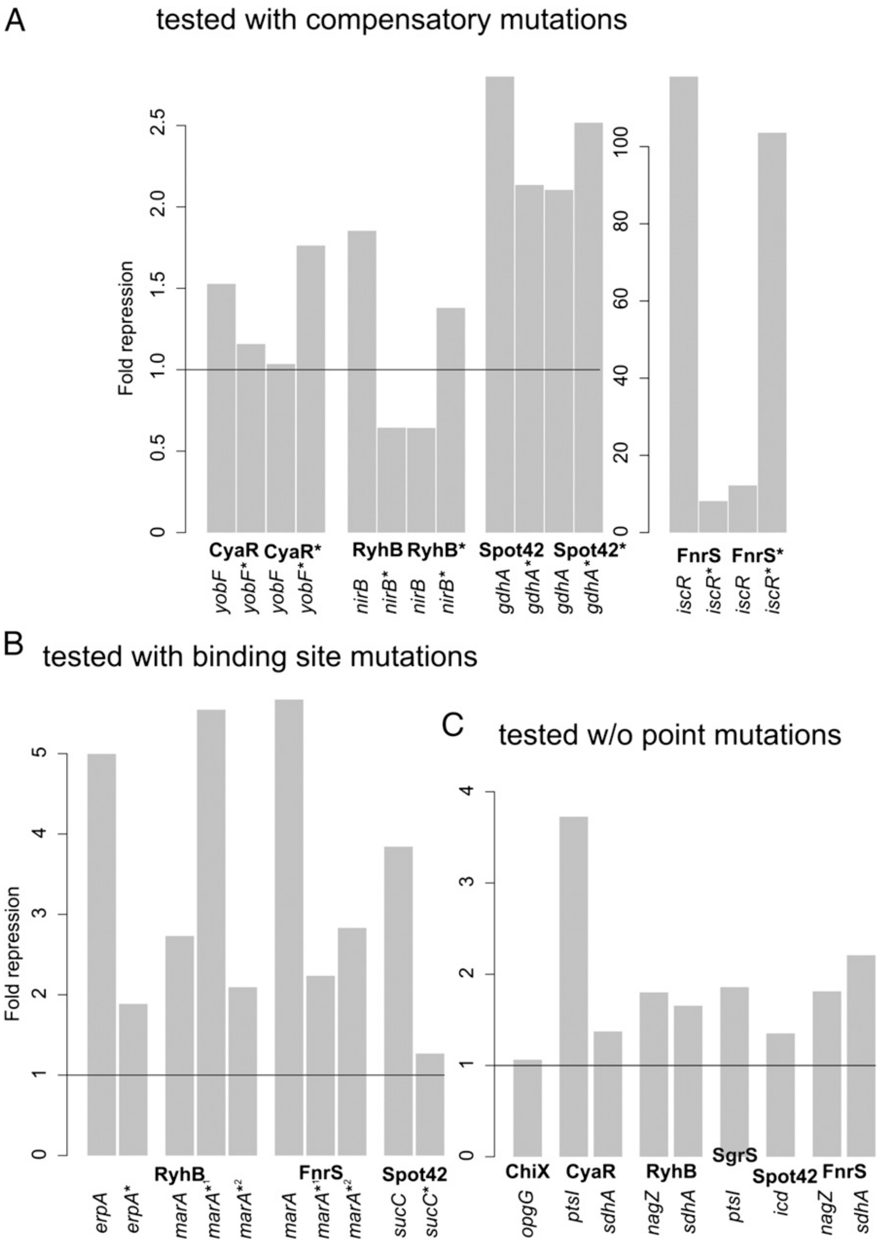


Fig. 4. Verification of sRNA target candidates. Translational repression of 5′ UTR–*gfp* fusions when overexpressing the sRNA. The fold repression is the ratio of the GFP fluorescence of the respective translational 5′ UTR–GFP fusion in the presence of the control plasmid pJV300 and a plasmid for the overexpression of the respective sRNA, after subtraction of the background fluorescence. Compensatory point mutations in the UTR and sRNA are indicated with an asterisk. (*A*) Verification of the *yobF*–CyaR, *nirB*–RyhB, *gdhA*–Spot42, and *iscR*–FnrS interactions with compensatory point mutations. (*B*) Verification of the *erpA*–RyhB, *marA*–RyhB, *marA*–FnrS, and *sucC*–Spot42 interactions with point mutations in the 5′UTR. (*C*) Verification of the *ptsI*–CyaR, *sdhA*–CyaR, *nagZ*–RyhB, *sdhA*–RyhB, *ptsI*–SgrS, *icd*–Spot42, *nagZ*–FnrS, and *sdhA*–FnrS interactions without point mutations.

FnrS, FsrA, GcvB, MicA, MicF, PrrF, RyhB, SgrS, and Spot42, bona fide physiological functions could be predicted accurately on our in silico results (Table 2). Compared with microarrays, CopraRNA has an advantage in that genetic modifications and time-consuming, expensive wet-laboratory experiments are not required for initial target screening. Additionally, CopraRNA is not biased by secondary effects, which might be picked up by experimental screening, and allows detection of targets not expressed under the tested conditions. Consequently, the predicted targets verify but also extend the existing microarray data.

However, CopraRNA also comes with certain limitations. The primary limitation of bioinformatic target prediction methods is that most predictions correspond to false positive predictions. The comparative approach of CopraRNA reduces this problem to the extent that further experimental analysis becomes much more reasonable than with existing tools, but it does not solve this problem completely. In our benchmark assay, half of the 101 known targets are detected with a *P* value threshold of 0.01 (*SI Appendix*, Fig S3*A*). At this threshold, an average of 65 targets is predicted for each sRNA and the FPR is ~95% (*SI Appendix*, Fig S3*B*). Thus, a reasonable sensitivity of 50% comes with a low specificity of 5%. In fact, this is a strong improvement, as the other tools tested reach a maximum sensitivity of 25% (IntaRNA) at 65 predictions per sRNA, and e.g., IntaRNA needs 226 predictions per sRNA to reach a sensitivity of 50%. Nevertheless, a low specificity challenges investigators to follow up on the predictions. For that reason, we do not stick to the *P* value threshold strictly, but focus on the top 15 list and on the predictions ($P \leq 0.01$) suggested by further postprocessing steps. These steps may include automatic and manual functional enrichment (Fig. 2), network analysis (Fig. 3), overlaps with transcription factor regulons (Fig. 5 and *SI Appendix*, Fig S13), or correlation patterns coming from microarray data (41, 42). This combined strategy was very successful in retaining sensitivity while enhancing specificity. We demonstrated this by the experimental verification of 73% of the selected 23 predicted targets that were not characterized previously. These results also show that the FPR is at least slightly overestimated because of previously unknown targets (*SI Appendix*, Fig S3*B*; compare dashed and solid blue lines). Another challenge is a prediction without a meaningful postprocessing result, caused, e.g., by the lack of additional data or lower prediction quality. For these cases, we control the FDR statistically by calculating a q-value. The average q-value at prediction rank 65 is ~0.54 and therefore judged by the current benchmark data, rather too optimistic. Nevertheless, the q-value distribution is valuable to roughly estimate the general prediction quality for a given sRNA. For example, we could not predict known targets for ArcZ. This less informative prediction is accompanied correctly by a rapidly growing q-value and only 10 predictions with $q \leq 0.5$. On the other side, the good prediction for GcvB has 38 predictions with $q \leq 0.5$, and as described above, the q-value fits well to the benchmark dataset. CopraRNA generally requires the conservation of an sRNA and also a substantial level of target conservation in the selected species. Therefore, single-organism–specific targets are likely to be missed, as are interactions that generally are not predictable by the underlying IntaRNA algorithm (e.g., double-kissing hairpin complexes). For example, the *metE*–FnrS interaction [verified in *E. coli* (20)] seems to be conserved or detectable only in three of the eight included species (*SI Appendix*, Fig. S12). This results in a high combined *P* value of 0.54 and a rank of 1,969 in the combined prediction and shows the importance of carefully selecting species. A small evolutionary distance favors sensitivity, and a large distance favors specificity. The downstream

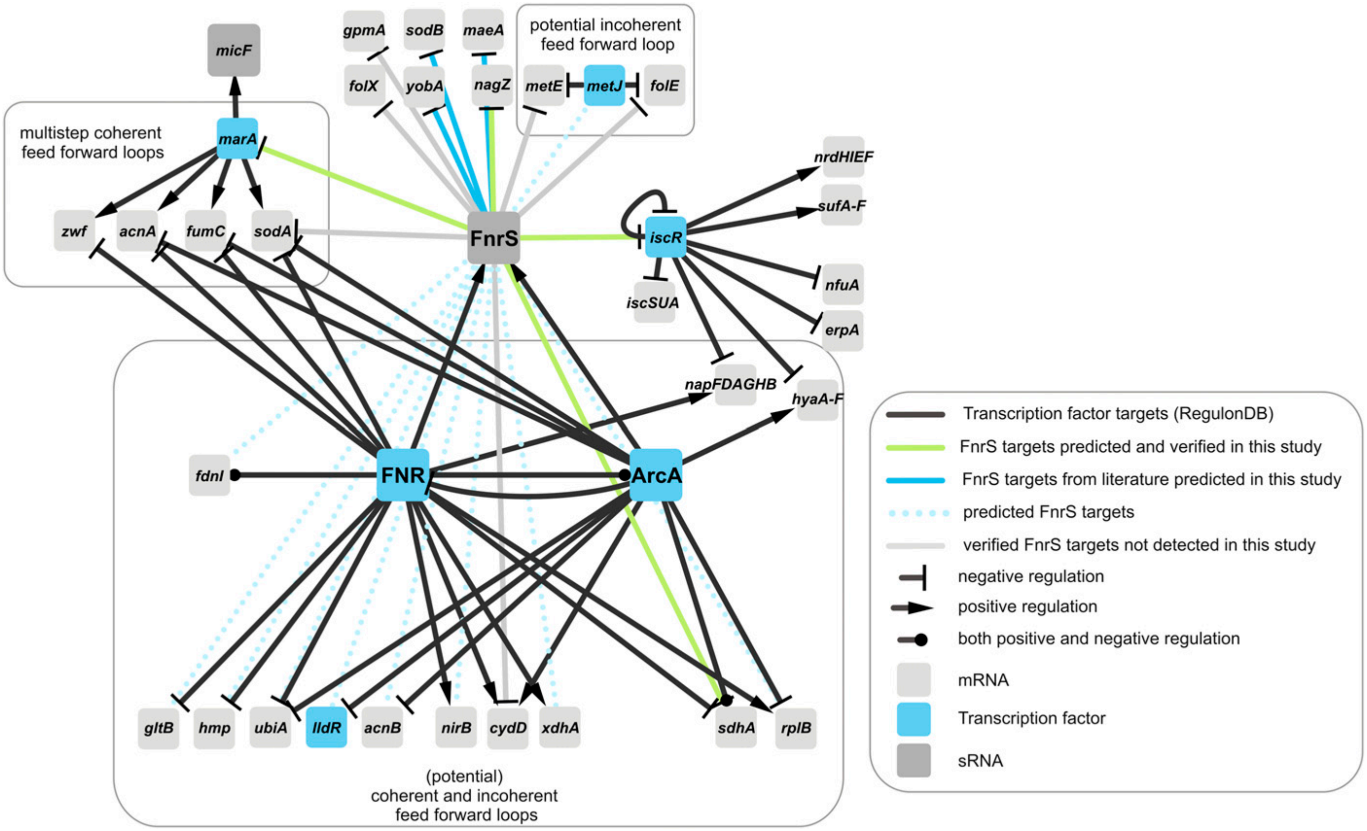

micF
gpmA
sodB
maeA
potential incoherent feed forward loop
folX
yobA
nagZ
metE
metJ
folE
multistep coherent feed forward loops
marA
zwf
acnA
fumC
sodA
FnrS
iscR
nrdHIEF
sufA-F
nfuA
erpA
iscSUA
napFDAGHB
hyaA-F
fdnI
FNR
ArcA
gltB
hmp
ubiA
lldR
acnB
nirB
cydD
xdhA
sdhA
rplB
(potential) coherent and incoherent feed forward loops
Transcription factor targets (RegulonDB)
FnrS targets predicted and verified in this study
FnrS targets from literature predicted in this study
predicted FnrS targets
verified FnrS targets not detected in this study
negative regulation
positive regulation
both positive and negative regulation
mRNA
Transcription factor
sRNA

Fig. 5. Partial regulatory network around FNR, ArcA, and FnrS. The figure shows verified FnrS targets, as well as predicted targets (CopraRNA *P* value ≤0.01) regulated by FNR or ArcA. For the transcription factors, only selected targets are displayed.

functional enrichment analysis relies on the availability of the organism in the DAVID database (31), and the results depend on the annotation quality of the genome of interest. Of note, CopraRNA is a target prediction tool for sRNAs that are expected to act in *trans*; it is not suitable for the differentiation of a *trans*-acting RNA from other types of transcripts. However, the functional enrichment analysis, the conservation plots, and the q-value distribution provided by CopraRNA might provide a hint as to whether a given conserved RNA is a functional *trans*-acting sRNA.

Additional Targets and Functions of Previously Characterized sRNAs. The inspection of the benchmark dataset revealed additional targets and functions, even for sRNAs extensively characterized in the past. For the cAMP receptor protein (CRP)-regulated sRNA CyaR (18, 19), we detected as yet unidentified targets in primary metabolism (*sdhA*) and the phosphotransferase system (*ptsI*), constituting previously unreported links of the CyaR regulon to carbon metabolism. Furthermore, with regard to the *yobF-cspC* operon, we found a potential explanation for the indirect negative effect of CyaR on the *rpoS* mRNA, which was detected in a screen with 26 sRNAs (43). The *yobF* gene is organized together with *cspC* in a dicistronic operon, and the RNA chaperone CspC is a posttranscriptional stabilizer of the *rpoS* message (44).

FnrS is involved in gene regulation after the shift from aerobic to anaerobic conditions, and its expression is activated by the transcription factors FNR and ArcA (20, 21). The combination of existing information (45) with our predictions and verifications for FnrS results in a remarkable complex regulatory network (Fig. 5): (*i*) FnrS transduces the signal to several non-FNR and -ArcA targets. These include the target *nagZ* and the two transcription factor mRNAs *iscR* and *marA*. (*ii*) The prediction also revealed several target candidates, which are controlled simultaneously by FNR and ArcA, which would establish multi-output feed-forward loops. Although the transcription factor MarA is not directly regulated by FNR or ArcA, four genes that are activated by MarA (*acnA*, *fumC*, *sodA*, *zwf*) are repressed by ArcA and/or FNR. These four genes are involved in the resistance to superoxide (46) and provide a reasonable explanation for the repression of *marA* by FnrS at anaerobic conditions. The repression of the transcription factor IscR may be part of the observed O_2-dependent expression of the *iscR* regulon (47).

FnrS shares three targets with RyhB. Both sRNAs regulate the mRNA encoding MarA, which is involved in the response to antimicrobial compounds and oxidative stress (46), and of the mRNA for the β-*N*-acetylglucosaminidase NagZ, which permits resistance to β-lactams in *Pseudomonas aeruginosa* (48). Interestingly, both MarA and NagZ are not obviously involved in iron homeostasis. For the iron stress-induced sRNA RyhB, we predicted mRNAs for 13 iron-containing proteins as targets and verified the posttranscriptional regulation of *erpA*, the mRNA of an A-type carrier (ATC) protein involved in iron–sulfur cluster biogenesis (49), and of *nirB*, which codes for a subunit of nitrite reductase.

Regarding the dual-function RNA SgrS, we predicted interactions with mRNAs of additional components of the phosphotransferase system (*chhB, cmtB* and *fruA*) and verified the posttranscriptional regulation of *ptsI* (Fig. 4), which codes for the non–sugar-specific enzyme I component of the PTS. Furthermore, we detected the recently described positive regulated sugar phosphatase mRNA *yigL* (50) as a direct target.

We also predicted and verified targets for the CRP-repressed Spot42 sRNA which is involved in catabolite repression and controls a range of genes in central and secondary metabolism and sugar transport (6). Our predictions show a large, 18-gene overlap with the CRP regulon and point to an even broader regulatory role for Spot42 in primary metabolism involving the citrate cycle and acetyl-CoA–dependent processes (Table 2, Tables S4 and S5, and *SI Appendix*, Fig. S13). Our successful experimental validation of the targets *gdhA*, *icd*, and *sucC* proves the accuracy of our predictions.

In sum, CopraRNA allows for an efficient screening of large numbers of sRNAs and has proven superior compared with existing methods. Using this tool, we obtained compelling evidence that sRNAs are global regulators of large sets of mRNAs, comparable to protein transcription factors and eukaryotic microRNAs. We also show that it is a common concept that mRNAs are targeted by multiple sRNAs and correctly predicted the regulatory hubs *csgD* and *rpoS*. Furthermore, we proposed and partially verified *gdhA*, *lrp*, *marA*, *nagZ*, *ptsI*, *sdhA*, and *yobF-cspC* as hubs targeted by up to seven different sRNAs. Finally, we present examples for complex posttranscriptional events at the operon level, including multiple targeting by the same, as well as different, sRNAs.

Methods

Experimental Methods. ***Bacterial strains and growth.*** Cells were grown in Luria–Bertani (LB) broth or on LB plates at 37 °C. Antibiotics (where appropriate) were applied at the following concentrations: 100 mg·mL^{-1} ampicillin and 25 mg·mL^{-1} chloramphenicol.

Plasmid construction. The plasmids for the overexpression of FnrS and CyaR and those for the translational superfolder–GFP fusions were constructed as described previously (22).

Oligonucleotides and plasmids. Oligonucleotides and plasmids are listed in *SI Appendix*, Tables S10 and S11.

Fluorescence measurements. Overnight cultures were used to inoculate (1:100) fresh cultures, and cultivation was continued to $OD_{600} = 2.0$. Culture samples equivalent to 1 OD were harvested by centrifugation and resuspended in PBS. Aliquots of 100 μL were transferred to a 96-well microtiter plate, and relative GFP levels were measured in a Victor3 fluorimeter (Perkin-Elmer). A wild-type strain was measured in parallel to subtract autofluorescence levels. All samples were measured in biological triplicates. This method was used to analyze the RyhB–*nirB* and the CyaR–*yobF* interactions.

Flow cytometry-based fluorescence measurements. Single bacterial colonies were inoculated in 200 μL LB medium in 96-well microtiter plates containing ampicillin and chloramphenicol and grown at 37 °C, 100 rpm for 12–15 h. Cells were diluted 1/5 in LB and fixed with formaldehyde (Roti-Histofix 10%; Carl Roth GmbH) to an final concentration of 1% (wt/vol) and measured directly on an Accuri C6 flow cytometer (BD Biosciences). The mean fluorescence of 50,000 events was averaged for 6–12 independent biological replicates. The fold repression was calculated as the ratio of the mean GFP fluorescence of the respective translational UTR–GFP fusion in the presence of the control plasmid pJV300 and a plasmid for the overexpression of the respective sRNA, after subtraction of the background fluorescence. Background fluorescence was measured with the control plasmids pXG-0 and pJV300 (22):

$$\mathrm{Fold}_{\mathrm{rep}} = \frac{\mathrm{Fluorescence\ UTR}_{\mathrm{pJV300}} - \mathrm{pXG}-0_{\mathrm{pJV300}}}{\mathrm{Fluorescence\ UTR}_{\mathrm{sRNA}} - \mathrm{pXG}-0_{\mathrm{pJV300}}}.$$

The respective mean fluorescences after subtraction of the background fluorescence are shown in *SI Appendix*, Fig. S9. Western blots were performed as described in ref. 9.

Theoretical Methods. ***Benchmark analysis.*** For the benchmark analysis, we conducted whole-genome target predictions for *E. coli* (NC_000913) and *S. enterica* (NC_003197, NC_003277) based on the sequences 200 nt upstream and 100 nt downstream of the annotated start codons as the input (the first nucleotide of the start codon corresponds to position 201). The Web server of RNApredator used the whole gene for target prediction. Otherwise, all the tools were used with the given standard parameters. The *P* value threshold of TargetRNA was set to 0.99 to obtain the top 100 predictions. The benchmark dataset included 18 sRNAs and a total of 101 previously published targets (*SI Appendix*, Table S1). Some targets were verified in both *E. coli* and *S. enterica*; the total number of verified sRNA–target pairs is 113, but we used only the nonredundant dataset. We included only targets for which a direct posttranscriptional regulation by an sRNA was verified experimentally. Targets detected only by RT-PCR, microarrays, or Northern blots and not verified further were excluded.

Functional enrichment. Functional enrichments (functional annotation clustering) were performed on the DAVID Web server (31) for all benchmark sRNA predictions. For each sRNA, the target candidates ($P \leq 0.01$) were tested against all the genes on the list as background. Obvious artifacts, i.e., predicted interactions with the complementary strand of the genomic coding region of the respective sRNA, were excluded. Enrichments were performed for *E. coli*. The standard parameters were changed to a "Similarity Threshold" of 0.85 and an "Initial Group Membership" and "Final Group Membership" of 2. Our threshold for a functional-enriched term was a DAVID enrichment score of ≥ 1.1. Networks were visualized using Cytoscape (51).

CopraRNA algorithm. To reduce the number of false positive hits in the interaction predictions, we searched for interactions that are conserved in various species. However, for several reasons, it is conceivable that the interaction is preserved whereas the actual interaction site is not. To be able to still predict conserved interactions, it is necessary to combine the evidence for interactions in the different species without resorting to a consensus-based approach. In addition to the Web server version, a stand-alone version of CopraRNA is available (www.bioinf.uni-freiburg.de/Software/). A more detailed description of CopraRNA, with a focus on the calculation of *P* values, may be found in *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported by the Deutsche Forschungsgemeinschaft Focus Program "Sensory and Regulatory RNAs in Prokaryotes" (SPP1258); Bundesministerium für Bildung und Forschung (BMBF) Grant 0316165 (to W.R.H, R.B., and J.V.); and the Excellence Initiative of the German Federal and State Governments (EXC 294 to R.B.). K.P. was supported by a postdoctoral fellowship from the Human Frontiers in Science Program.

1. Storz G, Vogel J, Wassarman KM (2011) Regulation by small RNAs in bacteria: Expanding frontiers. *Mol Cell* 43(6):880–891.
2. Gottesman S, Storz G (2011) Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* 3(12):a003798.
3. Papenfort K, Vogel J (2010) Regulatory RNA in bacterial pathogens. *Cell Host Microbe* 8(1):116–127.
4. Sharma CM, et al. (2011) Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* 81(5): 1144–1165.
5. Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA (2011) Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc Natl Acad Sci USA* 108(31):12875–12880.
6. Beisel CL, Storz G (2011) The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. *Mol Cell* 41(3):286–297.
7. Sharma CM, et al. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464(7286):250–255.
8. Mitschke J, et al. (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci USA* 108(5):2124–2129.
9. Papenfort K, Bouvier M, Mika F, Sharma CM, Vogel J (2010) Evidence for an autonomous 5′ target recognition domain in an Hfq-associated small RNA. *Proc Natl Acad Sci USA* 107(47):20435–20440.
10. Tjaden B, et al. (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res* 34(9):2791–2802.
11. Eggenhofer F, Tafer H, Stadler PF, Hofacker IL (2011) RNApredator: Fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res* 39(Web Server issue):W149–W154.
12. Busch A, Richter AS, Backofen R (2008) IntaRNA: Efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 24(24): 2849–2856.
13. Rehmsmeier M, Steffen P, Höchsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. *RNA* 10(10):1507–1517.
14. Hao Y, et al. (2011) Quantifying the sequence-function relation in gene silencing by bacterial small RNAs. *Proc Natl Acad Sci USA* 108(30):12473–12478.
15. Backofen R, Hess WR (2010) Computational prediction of sRNAs and their targets in bacteria. *RNA Biol* 7(1):33–42.
16. Richter AS, Backofen R (2012) Accessibility and conservation: General features of bacterial small RNA-mRNA interactions? *RNA Biol* 9(7):954–965.
17. Beisel CL, Updegrove TB, Janson BJ, Storz G (2012) Multiple factors dictate target selection by Hfq-binding small RNAs. *EMBO J* 31(8):1961–1974.
18. De Lay N, Gottesman S (2009) The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *J Bacteriol* 191(2):461–476.
19. Papenfort K, et al. (2008) Systematic deletion of *Salmonella* small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol Microbiol* 68(4):890–906.
20. Boysen A, Møller-Jensen J, Kallipolitis B, Valentin-Hansen P, Overgaard M (2010) Translational regulation of gene expression by an anaerobically induced small non-coding RNA in *Escherichia coli*. *J Biol Chem* 285(14):10690–10702.
21. Durand S, Storz G (2010) Reprogramming of anaerobic metabolism by the FnrS small RNA. *Mol Microbiol* 75(5):1215–1231.
22. Corcoran CP, et al. (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* 84(3):428–445.
23. Massé E, Vanderpool CK, Gottesman S (2005) Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol* 187(20):6962–6971.
24. Papenfort K, Podkaminski D, Hinton JCD, Vogel J (2012) The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. *Proc Natl Acad Sci USA* 109(13):E757–E764.
25. Uchiyama I (2007) MBGD: A platform for microbial comparative genomics based on the automated construction of orthologous groups. *Nucleic Acids Res* 35(Database issue):D343–D346.
26. Hartung J (1999) A note on combining dependent tests of significance. *Biom J* 41:849–855.
27. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100(16):9440–9445.
28. Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J (2008) Small RNA binding to 5′ mRNA coding region inhibits translational initiation. *Mol Cell* 32(6):827–837.
29. Argaman L, et al. (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 11(12):941–950.
30. Salvail H, Massé E (2012) Regulating iron storage and metabolism with RNA: An overview of posttranscriptional controls of intracellular iron homeostasis. *Wiley Interdiscip Rev RNA* 3(1):26–36.
31. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
32. Jørgensen MG, et al. (2012) Small regulatory RNAs control the multi-cellular adhesive lifestyle of *Escherichia coli*. *Mol Microbiol* 84(1):36–50.
33. Thomason MK, Fontaine F, De Lay N, Storz G (2012) A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol Microbiol* 84(1):17–35.
34. Holmqvist E, et al. (2010) Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J* 29(11):1840–1850.
35. Mika F, et al. (2012) Targeting of *csgD* by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in *Escherichia coli*. *Mol Microbiol* 84(1):51–65.
36. Holmqvist E, Unoson C, Reimegård J, Wagner EGH (2012) A mixed double negative feedback loop between the sRNA MicF and the global regulator Lrp. *Mol Microbiol* 84(3):414–427.
37. Desnoyers G, Massé E (2012) Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes Dev* 26(7):726–739.
38. Desnoyers G, Morissette A, Prévost K, Massé E (2009) Small RNA-induced differential degradation of the polycistronic mRNA *iscRSUA*. *EMBO J* 28(11):1551–1561.
39. Pfeiffer V, Papenfort K, Lucchini S, Hinton JCD, Vogel J (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol* 16(8):840–846.
40. Bandyra KJ, et al. (2012) The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Mol Cell* 47(6): 943–953.
41. Hernández-Prieto MA, et al. (2012) Iron deprivation in *Synechocystis*: Inference of pathways, non-coding RNAs, and regulatory elements from comprehensive expression profiling. *G3 (Bethesda)* 2(12):1475–1495.
42. Modi SR, Camacho DM, Kohanski MA, Walker GC, Collins JJ (2011) Functional characterization of bacterial sRNAs using a network biology approach. *Proc Natl Acad Sci USA* 108(37):15522–15527.
43. Mandin P, Gottesman S (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO J* 29(18):3094–3107.
44. Cohen-Or I, Shenhar Y, Biran D, Ron EZ (2010) CspC regulates *rpoS* transcript levels and complements *hfq* deletions. *Res Microbiol* 161(8):694–700.
45. Gama-Castro S, et al. (2011) RegulonDB version 7.0: Transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res* 39(Database issue):D98–D105.
46. Martin RG, Rosner JL (2011) Promoter discrimination at class I MarA regulon promoters mediated by glutamic acid 89 of the MarA transcriptional activator of *Escherichia coli*. *J Bacteriol* 193(2):506–515.
47. Giel JL, Rodionov D, Liu M, Blattner FR, Kiley PJ (2006) IscR-dependent gene expression links iron-sulphur cluster assembly to the control of O_2-regulated genes in *Escherichia coli*. *Mol Microbiol* 60(4):1058–1075.
48. Zamorano L, et al. (2010) NagZ inactivation prevents and reverts β-lactam resistance, driven by AmpD and PBP 4 mutations, in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 54(9):3557–3563.
49. Pinske C, Sawers RG (2012) A-type carrier protein ErpA is essential for formation of an active formate-nitrate respiratory pathway in *Escherichia coli* K-12. *J Bacteriol* 194(2): 346–353.
50. Papenfort K, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J (2013) Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* 153(2): 426–437.
51. Cline MS, et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2(10):2366–2382.
52. Keseler IM, et al. (2013) EcoCyc: Fusing model organism databases with systems biology. *Nucleic Acids Res* 41(Database issue):D605–D612.

5 Two separate modules of the conserved regulatory RNA AbcR1 address multiple target mRNAs in and outside of the translation initiation region

Aaron Overlöper, Alexander Kraus, Rosemarie Gurski, **Patrick R. Wright**, Jens Georg, Wolfgang R. Hess and Franz Narberhaus (2014) **RNA Biology**, 11, 624-640.
DOI: 10.4161/rna.29145

Personal contribution

In this project, I was involved in performing the CopraRNA in silico analyses for the small RNA AbcR1.

Patrick R. Wright

The following co-authors confirm the above stated contribution.

Prof. Dr. Franz Narberhaus

Dr. Aaron Overlöper

RNA Biology 11:5, 624–640; May 2014; © 2014 Landes Bioscience

Two separate modules of the conserved regulatory RNA AbcR1 address multiple target mRNAs in and outside of the translation initiation region

Aaron Overlöper[1], Alexander Kraus[1], Rosemarie Gurski[1], Patrick R Wright[2,†], Jens Georg[2], Wolfgang R Hess[2], and Franz Narberhaus[1]*

[1]Microbial Biology; Ruhr University Bochum; Germany; [2]Genetics and Experimental Bioinformatics; University of Freiburg; Germany

[†]Current affiliation: Bioinformatics Group; Department of Computer Science; University of Freiburg; Germany

Keywords: regulatory RNA, small RNA, ABC transporter, RNA-RNA interaction, alpha-proteobacteria, Agrobacterium

The small RNA AbcR1 regulates the expression of ABC transporters in the plant pathogen *Agrobacterium tumefaciens*, the plant symbiont *Sinorhizobium meliloti*, and the human pathogen *Brucella abortus*. A combination of proteomic and bioinformatic approaches suggested dozens of AbcR1 targets in *A. tumefaciens*. Several of these newly discovered targets are involved in the uptake of amino acids, their derivatives, and sugars. Among the latter is the periplasmic sugar-binding protein ChvE, a component of the virulence signal transduction system. We examined 16 targets and their interaction with AbcR1 in close detail. In addition to the previously described mRNA interaction site of AbcR1 (M1), the CopraRNA program predicted a second functional module (M2) as target-binding site. Both M1 and M2 contain single-stranded anti-SD motifs. Using mutated AbcR1 variants, we systematically tested by band shift experiments, which sRNA region is responsible for mRNA binding and gene regulation. On the target site, we find that AbcR1 interacts with some mRNAs in the translation initiation region and with others far into their coding sequence. Our data show that AbcR1 is a versatile master regulator of nutrient uptake systems in *A. tumefaciens* and related bacteria.

Introduction

Within the last decade it has become increasingly clear that small RNAs (sRNAs) are equally efficient and versatile regulators of gene expression as protein-based transcription factors. Most *trans*-encoded sRNAs act at the post-transcriptional level by base-pairing to target mRNAs and can have a positive or negative effect on gene expression by affecting translation and/or RNA decay.[1,2] Small RNAs typically offer only limited complementarity to their targets. A segment of only seven contiguous bases, the so-called seed region, can be sufficient to confer specificity.[3-5] Therefore, sRNAs are well suited for regulation of multiple mRNAs. Another level of complexity is reached when a single mRNA is subject to regulation by several sRNAs.[6] Overall, this can lead to large sRNA-based regulatory networks that sense and respond to the nutritional status of the cell.[7,8]

The fundamental importance of sRNAs is reflected by their involvement in numerous cellular processes, like cell division (DicF), transcription (6S RNA), photosynthesis (PcrZ), stress adaption (OxyS), virulence, quorum sensing (Qrr), carbon storage (CsrBC), and phosphosugar metabolism.[9-21] A class of genes frequently controlled by sRNAs codes for periplasmic substrate binding proteins of bacterial ABC transporters.[7,22-26] This transporter superfamily uses periplasmic solute-binding proteins to take up a wide range of substrates (sugars, amino acids and their derivatives, as well as proteins and drugs).[27-29]

Most of our knowledge on sRNAs derives from studies with *Escherichia coli* and *Salmonella*. However, deep sequencing-assisted approaches have revealed numerous sRNAs in any given bacterium or archaeon.[30] Experimental evidence for a regulatory function of these small-sized RNAs has been provided in a limited number of cases, for example, in *Bacillus subtilis* and other Gram-positives, in cyanobacteria, archaea, *Rhodobacter*, and *Xanthomonas*.[11,31-35]

Genome-wide surveys have recently revealed hundreds of sRNAs in the plant pathogen *Agrobacterium tumefaciens*.[36,37] This bacterium is able to induce tumors (crown galls) upon transfer of a DNA fragment (T-DNA) from its tumor-inducing (Ti) plasmid to the nuclear genome of the host plant.[38,39] In the transformed plant cells, expression of T-DNA encoded growth factor genes

*Correspondence to: Franz Narberhaus; Email: franz.narberhaus@rub.de
Submitted: 03/07/2014; Revised: 05/02/2014; Accepted: 05/07/2014; Published Online: 05/13/2014
http://dx.doi.org/10.4161/rna.29145

results in cell proliferation and tumor formation. Additionally, plant metabolism is re-programmed to produce opines serving as carbon and nitrogen source for *A. tumefaciens*.[40,41] Perception of plant-derived signals involves several bacterial factors, including the two-component system VirA/VirG, which mediates the activation of the virulence cascade, and ChvE, a periplasmic substrate binding protein that binds host-derived sugars and plays a role in activation of the virulence cascade (ChvE).[42-45] Other putative substrate-binding proteins are involved in attachment (AttC), host defense (Atu2422 and Atu4243), and agrocinopine uptake.[46-49] In addition to these specialized functions in plant-microbe interaction, ABC transport systems are required for regular nutrient acquisition in *A. tumefaciens* like in other free-living bacteria.[50]

At least three ABC transporters in *A. tumefaciens* are under negative control of the sRNA AbcR1 (ABC transporter regulator 1).[26] Among the targets is *atu2422* encoding the binding protein for GABA (γ-amino butyric acid), a plant-derived defense molecule that interferes with quorum sensing in *Agrobacterium*.[47,51] AbcR1 is encoded in an intergenic region in tandem with the related sRNA AbcR2.[26] Both are maximally expressed in the late exponential phase. Currently, there is no evidence that AbcR2 plays a regulatory role in *A. tumefaciens*.

Like *Agrobacterium*, various *Rhizobium* species encode numerous sRNAs, including homologs of AbcR1 and AbcR2.[52-56] In contrast to *Agrobacterium*, AbcR1 and AbcR2 in *Sinorhizobium* are divergently expressed, namely the first is present in exponential phase whereas the second accumulates in stationary phase suggesting that they operate at different conditions.[57] The amino acid binding protein LivK was found to be controlled by AbcR1 but not AbcR2.[57] Two other ABC transporter genes are negatively regulated by AbcR1 and AbcR2.[58] In *Brucella abortus*, both AbcR1 and AbcR2 seem to have at least some redundant function.[59] Microarray analysis revealed about 25 elevated transcripts, several coding for ABC transporters, in the double mutant. At least three of these transcripts can be controlled by AbcR1 or AbcR2 alone. Moreover, only the double mutant but neither single mutant was attenuated in macrophages and in mice. The commonalities and differences in AbcR1-mediated gene regulation in these model organisms certainly warrants further studies to understand the role of this conserved sRNA in a plant pathogen, a plant symbiont, and a human pathogen.

AbcR1 belongs to the large group of Hfq-associated sRNAs.[55,59-61] Hfq is an RNA chaperone that facilitates base-pairing between sRNAs and their targets.[62,63] About 10 ABC transporter proteins were found to accumulate in an *A. tumefaciens* Δ*hfq* mutant and we wondered whether more than the previously identified three targets *atu2422*, *atu1879*, and *frcC* were controlled by AbcR1.[26,61] We used a combination of proteomics and bioinformatics approaches to identify numerous new targets of AbcR1. RNA–RNA interactions studies revealed that AbcR1 uses two separate regions to address mRNAs either in the translation initiation region (TIR) or far downstream in the coding region. Our results support the function of AbcR1 as versatile master regulator to control *Agrobacterium* physiology.

Results

AbcR1 regulates periplasmic binding proteins of several ABC transporters

To identify new targets of AbcR1, we compared the proteomes of the marker-less AbcR1 mutant (ΔAbcR1) and the wild-type (WT) strain by two-dimensional PAGE. Cultures were grown to stationary phase (OD600 of 1.5) when AbcR1 is maximally expressed.[26] Total protein extracts from three biological replicates were subjected to two-dimensional PAGE and the relative protein abundance was visualized by dual-channel images (**Fig. S1**). Proteins equally abundant in the WT and mutant appear as yellow spots, whereas proteins overrepresented in WT or ΔAbcR1 are green or red, respectively. Overall, 68 proteins were affected by the presence of AbcR1, indicating potential targets of AbcR1 (**Table S1**). Twenty-five were up and 43 downregulated. Twenty candidates were extracted from the gel, digested with trypsin, and subjected to mass spectrometry (**Table 1**). The presence of the known targets Atu2422 and Atu1879 among them validated this approach. Northern blot experiments revealed that the increased protein levels in the ΔAbcR1 mutant correlated with increased mRNA levels of *atu2422* and *atu1879* in stationary phase (**Fig. 1A and B**).

Validation of eight new AbcR1 targets

The 18 other AbcR1-dependent proteins were so-far-unknown candidates (**Table S1**). To recapitulate AbcR1-mediated regulation at the mRNA level, eight of the new candidates were chosen for northern blot analysis with *Agrobacterium* WT and ΔAbcR1 mutant grown to exponential (OD600 of 0.5) and stationary (OD600 of 1.5) phase. The mRNAs of five periplasmic binding proteins of ABC transporters (Atu4577, MalE, Atu4046, Atu4678, and DppA) showed clear AbcR1-dependent regulation consistent with elevated protein levels in the ΔAbcR1 strain (**Fig. 2A–E**). The same was true for Atu0857, an annotated oxidoreductase (**Fig. 2F**). The *frcB* transcript appears to be downregulated by AbcR1 in the exponential growth phase but, consistent with 2D PAGE, upregulated in stationary phase (**Fig. 2G**). Reduced transcript levels of *atpH* in ΔAbcR1 in exponential growth supported positive regulation by AbcR1 as seen on the protein level (**Fig. 2H**). Transcripts of *atpH* and *dppA* (**Fig. 2B**) were only detectable in exponential growth phase suggesting that they undergo a rapid turnover in later growth phases.

Overlap between AbcR1- and Hfq-dependent mRNAs

The AbcR1-dependent genes *malE*, *atu4678*, and *dppA* have recently been shown to be affected by Hfq.[61] In that study, several proteins overrepresented in the *A. tumefaciens hfq* mutant were isolated from 1D SDS-PAGE gels and identified as ABC transporters. This led us to assume that a more comprehensive profile of the Δ*hfq* proteome might reveal additional AbcR1 targets. Upon separation by 2D-SDS-PAGE, 31 putative Hfq-dependent proteins were selected and identified by mass spectrometry (**Fig. 3A**). Among them were many periplasmic binding proteins of ABC transporters (**Table S2**) and 10 proteins identified as AbcR1 targets were also affected by the absence of Hfq (**Fig. 3B**) indicating that AbcR1 acts through Hfq as previously shown for the target Atu2422.[61]

Table 1. Potential AbcR1 targets in *A. tumefaciens* identified by 2D proteomics

Protein	Locus tag	Product	ΔAbcR1/WT
Atu4577	*atu4577*	ABC transporter substrate binding protein	66,90
PykA	*atu3762*	pyruvate kinase	36,86
RbsB	*atu3821*	ABC transporter substrate-binding protein (ribose)	25,64
Atu0857	*atu0857*	oxidoreductase	13,45
Atu2188	*atu2188*	oxidoreductase	9,42
MalE	*atu2601*	ABC transporter, substrate binding protein (maltose)	8,05
Pgi	*atu0404*	glucose-6-phosphate isomerase	6,64
Atu4046	*atu4046*	ABC transporter substrate-binding protein (glycine betaine)	6,22
MurE	*atu2099*	UDP-N-acetylmuramoylalanyl-D-glutamate-2,6- diaminopimelate ligase	5,87
Atu1879	*atu1879*	ABC transporter, substrate binding protein (amino acid)	4,50
Atu0157	*atu0157*	ABC transporter, substrate binding protein	3,87
Atu4678	*atu4678*	ABC transporter substrate-binding protein (amino acid)	3,85
Atu2422	*atu2422*	ABC transporter, substrate binding protein (amino acid GABA)	3,66
FrcB	*atu0063*	ABC transporter, substrate binding protein (sugar)	2,13
DppA	*atu4113*	ABC transporter substrate-binding protein (dipeptide)	2,07
Atu3259	*atu3259*	dehydrogenase	0,20
RplI	*atu1088*	50S ribosomal protein L9	0,18
AtpH	*atu2625*	ATP Synthase delta chain	0,16
MurB	*atu2092*	UDP-N-acetylenolpyruvoylglucosamine reductase	0,08
RplY	*atu2227*	50S ribosomal protein L25	0,05

List of proteins with altered abundance in three replicates of the ΔAbcR1 strain in comparison to the WT (fold changes < 0.5 or > 2, respectively). Quantitative proteomics was performed by two-dimensional PAGE with total protein samples from stationary growth phase (OD600: 1.5) of the *A. tumefaciens* wild-type (WT) and the AbcR1 deletion mutant (ΔAbcR1) followed by MALDI-TOF analysis. The entire list of all proteins significantly accumulated in WT or in ΔAbcR1 can be found in **Figure S1**.

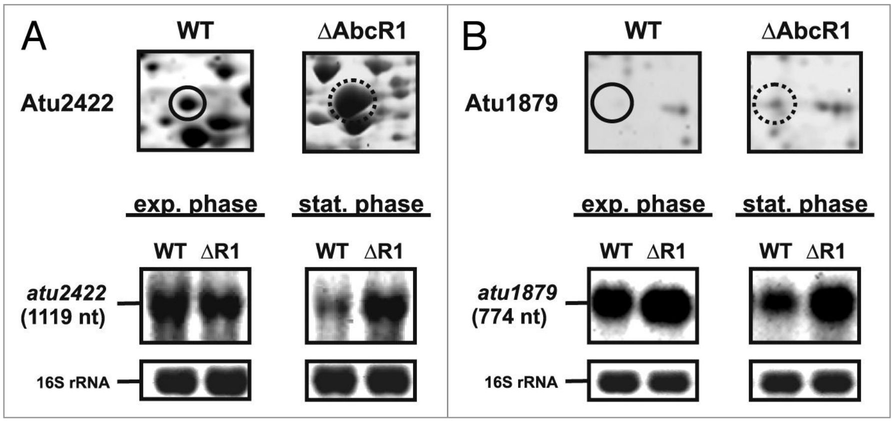


Figure 1. Identification of known AbcR1 targets by 2D-PAGE. Subsections of 2D gels showing Atu2422 (**A**) and Atu1879 (**B**) from *A. tumefaciens* WT (closed black circle) and ΔAbcR1 deletion mutant (dotted black circle) and northern blot analyses of *atu2422* (**B**) and *atu1879* (**C**) transcripts in different growth phases. The WT and the ΔAbcR1 deletion mutant (ΔR1) were grown to exponential (OD600: 0.5) or stationary phase (OD600: 1.5) in YEB medium. Eight µg of total RNA were separated on 1.2% denaturing agarose gels. Ethidiumbromide-stained 16S rRNAs were used as loading control.

Validation of six more AbcR1 targets

Northern blot experiments with probes against the two Hfq targets *atu4431* (**Fig. 3**) and *atu4259* confirmed regulation by AbcR1 as their mRNAs accumulated in the sRNA mutant (**Fig. 4A and B**; note that migration of the very abundant 16S rRNA to a similar position in the gel interferes with detection of the mRNAs and results in two bands).[61] One particularly interesting protein affected by Hfq was ChvE, a periplasmic sugar-binding protein involved in host sensing of *A. tumefaciens*.[43,45,64] Its regulatory pattern resembles that of FrcB. Both proteins were less abundant in the *hfq* deletion strain than in the WT (**Fig. 3A**; **Table S2**). In contrast to most other AbcR1 targets, but like the *frcB* transcript (**Fig. 2G**), the *chvE* mRNA was slightly downregulated in the absence of AbcR1 in exponential phase but clearly upregulated in stationary phase (**Fig. 4C**) suggesting growth phase-dependent regulation by AbcR1. Regulation of ChvE by AbcR1 raised our interest in NocT and AttC, substrate binding proteins of putative virulence-related ABC transporters required for the uptake of plant-synthesized nopaline (NocT) or for the transport spermidine and putrescine (AttC).[46] They were not detected by 2D PAGE analysis. However, northern blot analysis revealed that they clearly are AbcR1 targets. *nocT* is a typical negatively controlled AbcR1 target (**Fig. 4D**) whereas regulation of *attC* varies depending on the growth condition (**Fig. 4E**). The final potential AbcR1 target was predicted by the CopraRNA algorithm (Comparative Prediction Algorithm for sRNA Targets, see below).[8] Atu3114 was not identified by our proteomics approaches but northern blot analysis showed AbcR1-dependent regulation (**Fig. 4F**).

CopraRNA predicts two functional AbcR1 modules and variable target-binding regions

Having identified at least 16 AbcR1-dependent genes, we wondered whether they are all regulated by base pairing of the RBS with the first exposed loop of AbcR1 as documented for *atu2422* and *S. meliloti livK*.[26,57] To computationally predict interaction regions between AbcR1 and its target mRNAs, we made use of the recently established CopraRNA program.[8] It integrates phylogenetic information to predict sRNA–mRNA interactions on the genomic scale. An alignment of orthologous AbcR1 sequences from *A. tumefaciens* C58, *Agrobacterium radiobacter* K84, *Rhizobium etli* CFN42, *Rhizobium leguminosarum* bv. vicae, *Rhizobium etli* 652, *Sinorhizobium meliloti* 1021, and *Sinorhizobium medicae* WSM419 revealed long almost identical sequence stretches (**Fig. 5A**). The secondary structures were compared using the ClustalW2 program prior to calculation of a consensus structure with the RNAalifold webserver.[65,66] Regions highly conserved in sequence are equally conserved in structure (**Fig. 5B**). Like the experimentally mapped structure of *A. tumefaciens* AbcR1, the sRNAs fold into three hairpins.[26] Apart from the *atu2422* interaction site (module 1 = M1), a second conserved single-stranded region (M2) was found between the first and second hairpin. Both regions contain a UCCC motif potentially able to interact with SD-like sequences (**Fig. 5A and B**). A domain prediction of putatively interacting sites between AbcR1 and 15 of the target mRNAs validated in this study

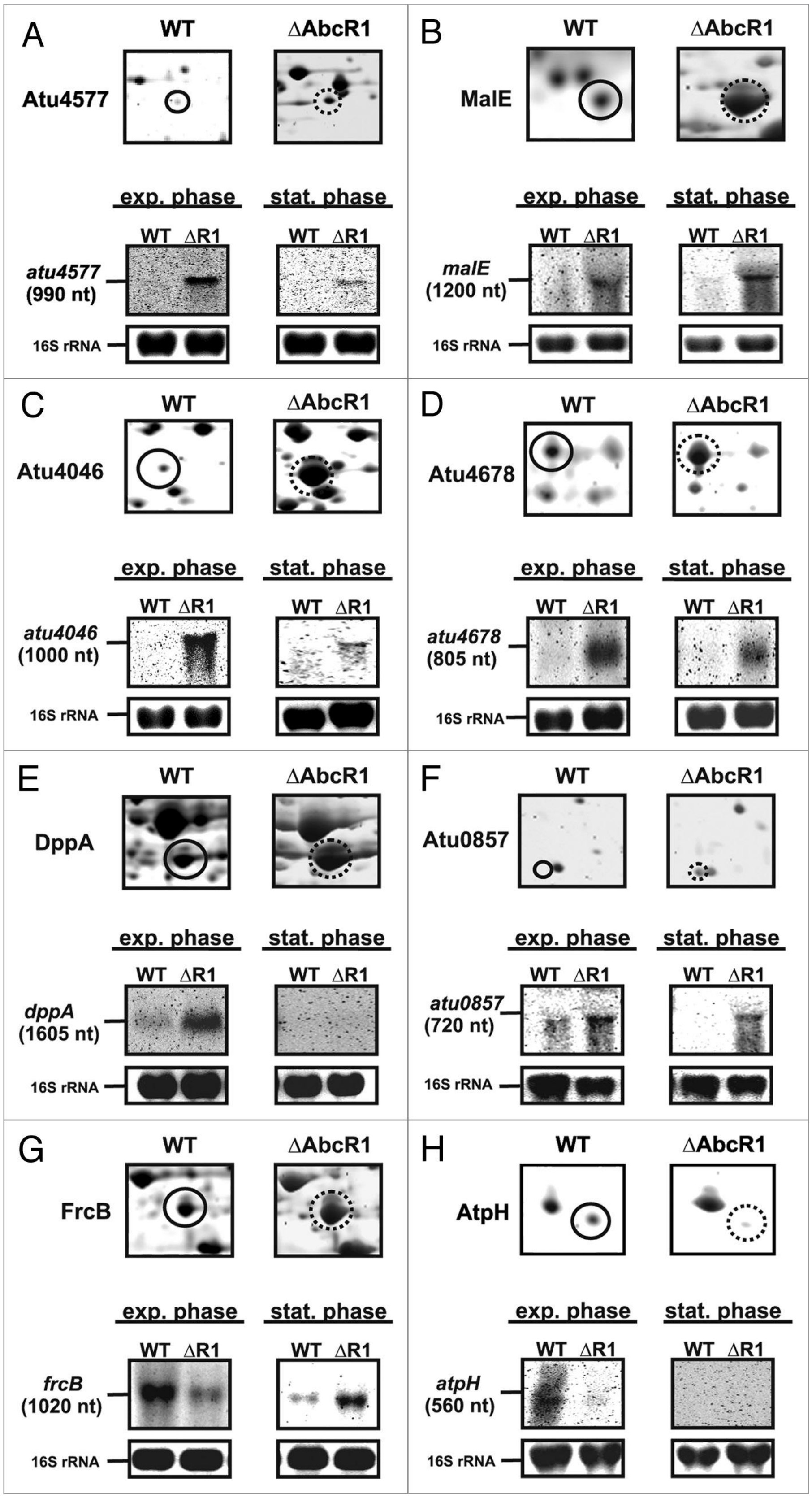


Figure 2. Validation of new AbcR1 targets. Subsections of 2D gels showing Atu4577 (**A**), MalE (**B**), Atu4046 (**C**), Atu4678 (**D**), DppA (**E**), Atu0857 (**F**), FrcB (**G**), and AtpH (**H**) from *A. tumefaciens* wild-type (closed black circles) and ΔAbcR1 mutant (dotted black circles) and corresponding northern blot analyses of target mRNAs in different growth phases. The wild-type (WT) and the ΔAbcR1 deletion mutant (ΔR1) were grown and treated as in **Figure 1**.

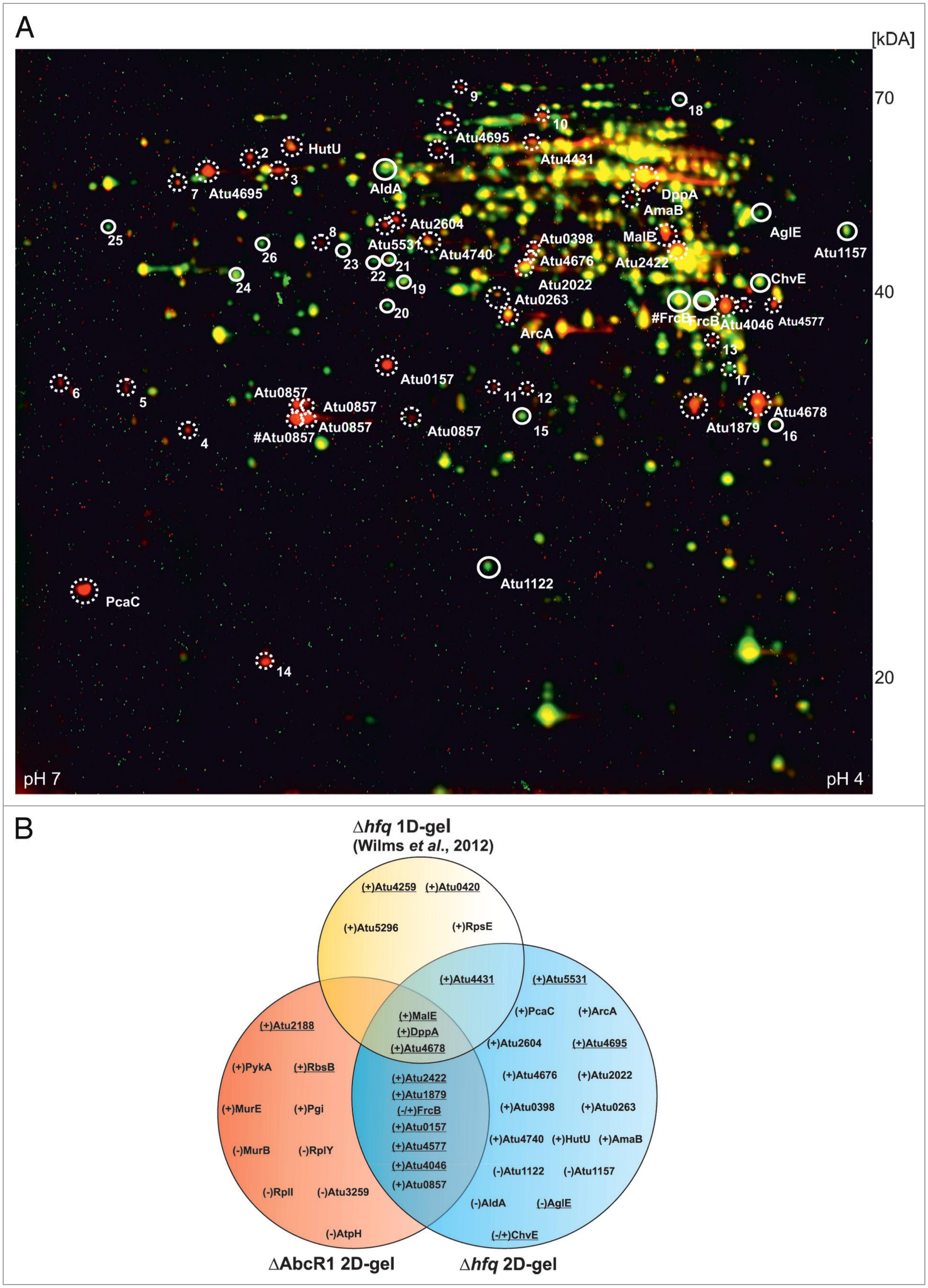


Figure 3. For figure legend, see page 629.

Figure 3 (See opposite page page). Altered protein synthesis in the Δ*hfq* mutant reveals new putative AbcR1 targets. (**A**) Total protein samples from stationary growth phase (OD600: 1.0) of *A. tumefaciens* WT and the *hfq* deletion mutant (Δ*hfq)* were subjected to two-dimensional gel electrophoresis. Proteins overrepresented in WT or in Δ*hfq* preparations are shown in green or red, respectively. Proteins present in equal amounts in both preparations appear in yellow. Six (green spots) and 22 (red spots) proteins significantly accumulated equally in three replicates of WT (closed white circles) or in Δ*hfq* (dotted white circles) were analyzed via MALDI-TOF. Altered proteins that were not identified by MALDI-TOF are marked with numbers. #, proteins were identified in multiple spots. (**B**) Venn diagram comparing altered proteins from different proteomic approaches (Δ*hfq* 1D-gel, Δ*hfq* 2D-gel, and ΔAbcR1 2D-gel) in *A. tumefaciens*. ABC transporter components are underlined. (+) or (-) indicate over- or underrepresentation of proteins in deletion mutants.

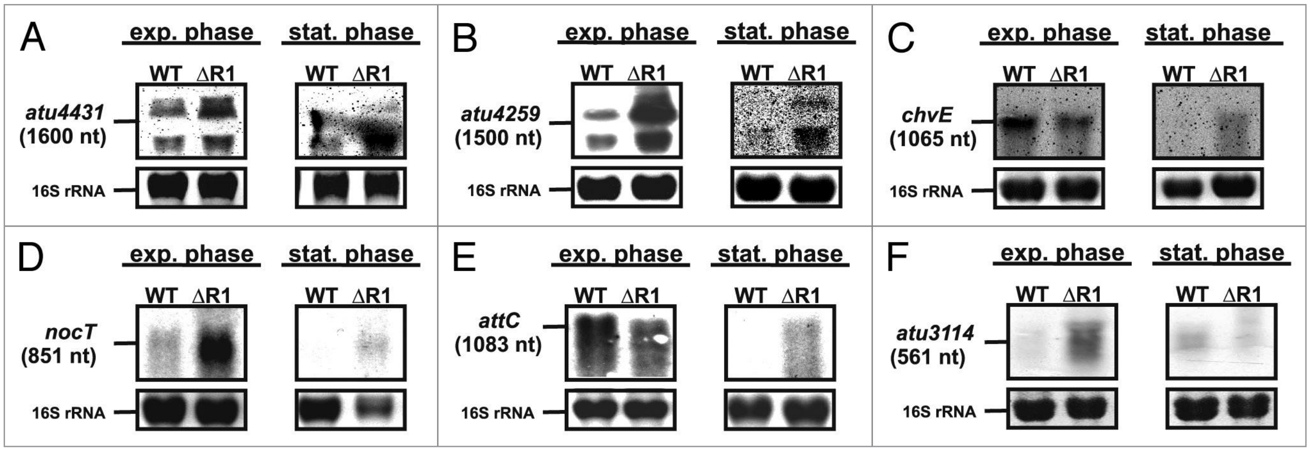


Figure 4. Verification of additional AbcR1 targets. Northern blot analyses of *atu4413* (**A**), *atu4259* (**B**), *chvE* (**C**), *nocT* (**D**), *attC* (**E**), and *atu3114* (**F**) mRNAs in different growth phases. The wild-type and the ΔAbcR1 deletion mutant (ΔR1) were grown to exponential (OD600: 0.5) or stationary phase (OD600: 1.5) in YEB medium. Eight µg of total RNA were separated on 1.2% denaturing agarose gels. Ethidiumbromide-stained 16S rRNAs were used as loading control.

(note that *atu4577* could not be used because it is not a conserved gene) suggested that both M1 and M2 are involved in target recognition (**Fig. 5C**, for visualization of detailed AbcR1 M1- and M2-target mRNA interactions in *A. tumefaciens* see **Table S3**). The predicted mRNA interaction sites preferentially lie around the SD sequence but several sites are located far into the coding region (**Fig. 5D**; **Table S3**).

AbcR1 discriminates between target mRNAs through two target-binding regions

A series of band shift experiments was used to experimentally validate the algorithmically predicted RNA–RNA interactions. Four different in vitro synthesized AbcR1 RNAs were used: the WT RNA, Mut1 with a UCC-AAA exchange in M1, Mut2 with a UCCC-AAAA exchange in M2, and the combined Mut1+2 exchange (**Fig. 6A**). Band shift experiments were performed to verify interactions between AbcR1 variants and their target mRNAs. In the first round, the ^{32}P-labeled AbcR1 variants were incubated with increasing concentrations of four different targets predicted to be addressed around the TIR. The target RNAs consisted of 100 to 150 nucleotides containing the predicted interaction region. As expected, band shifts with AbcR1 and AbcR1 Mut2 but not with the Mut1 and Mut1+2 RNAs confirmed complex formation between the TIRs of *atu2422* and *frcB* with AbcR1 region M1 (**Fig. 6B**). Conversely, the *atu4678* and *chvE* TIRs were shown to interact with AbcR1 region M2 (**Fig. 6C**).

Underrepresentation of AtpH protein and *atpH* mRNA in the absence of AbcR1 (**Fig. 2H**) suggested positive regulation by the sRNA. CopraRNA predicted an interaction between the TIR of *atpH* and AbcR1 region M2 including the adjacent hairpin 2 (**Figs. 5C and D**). In agreement with this prediction, mutations in M1 or M2 alone and even in both M1 and M2 were not sufficient to fully abrogate AbcR1–*atpH* interaction suggesting an extended interaction (**Fig. 6D**). The sRNA–mRNA interaction was lost in the simultaneous presence of a mutation in M2 of AbcR1 and in the *atpH* TIR (*atpH*-5U; **Fig. 6D**).

In a second round of experiments, three mRNAs predicted to interact with AbcR1 in their CDS were tested. One hundred and fifty nt-long RNAs containing the predicted interaction region were incubated with the radiolabeled AbcR1 variants. Interaction of M1-containing AbcR1 in the CDS of *atu1879* (**Fig. 7A**) explains why the TIR of *atu1879* could not be shifted in our previous study.[26] Region M1 also interacts with the CDS of *atu3114*. Extending the CopraRNA prediction, the CDS region of *malE* was not only able to interact with module M1 but also retarded the M2 RNA (**Fig. 7B**) suggesting that both modules are able to initiate seed pairing. Consistent with this assumption, the Mut1+2 RNA was unable to shift the *malE* fragment.

In vivo verification of target binding by AbcR1 modules M1 and M2

To validate the in vitro results on the interaction of AbcR1 with its target mRNAs in vivo, we used an *A. tumefaciens* ΔAbcR1/2 double mutant complemented with the empty vector (+v in **Fig. 8**) or a plasmid constitutively expressing one of the four AbcR1 variants (+AbcR1, +Mut1, +Mut2, or +Mut1+2). Production of the AbcR1 transcripts was confirmed by northern blot analysis (**Fig. 8A**). The mRNA levels of four different AbcR1 targets were determined by northern blot analysis. Consistent with the band shift experiments (**Figs. 6 and** 7), region M1 was

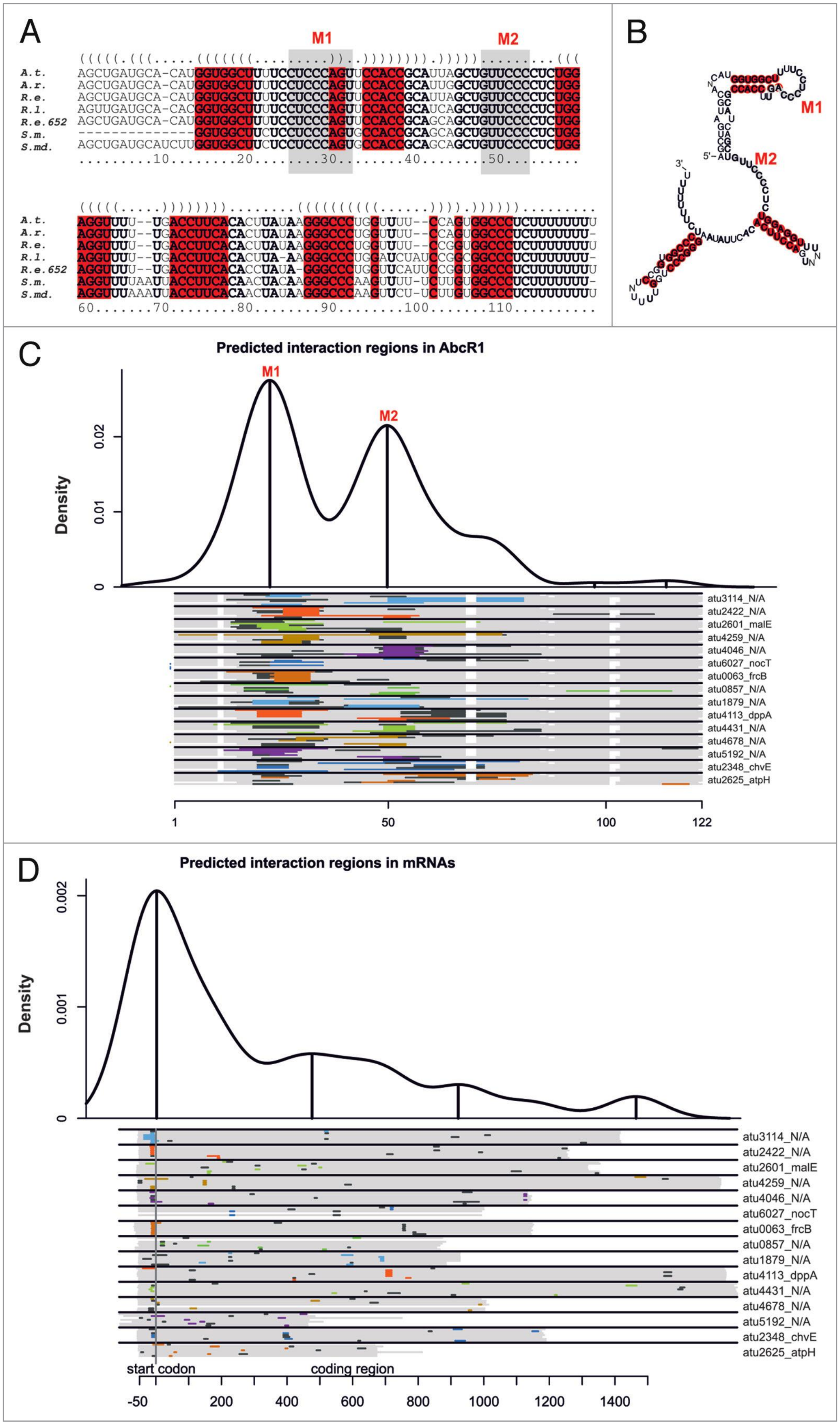


Figure 5. For figure legend, see page 631.

Figure 5 (see opposite page). Comparative computational predictions suggest two functional AbcR1 modules. Sequence alignment (**A**) and consensus structure (**B**) of AbcR1 in *A. tumefaciens* C58 (*A.t.*), *A. radiobacter* K84 (*A.r.*), *R. etli* CFN42 (*R.e.*), *R. leguminosarum* bv. vicae (*R.l.*), *R. etli* 652 (*R.e.*652) *S. meliloti* 1021 (*S.m.*), and *S. medicae* WSM419 (*S.md.*). Sequence conservation is given in bold letters. Nucleotides highly conserved in structure are marked in red. The calculated structure is given in dot-bracket notation above the alignment. Grey shaded boxes (**A**) or gray marked nucleotides (**B**) represent regions M1 and M2. Visualization of the predicted binding regions in AbcR1 (**C**) and the experimentally verified targets of AbcR1 (**D**). The density plots at the top indicate the relative frequency of sRNA or mRNA nucleotide positions in the predicted AbcR1–target mRNA interactions. The density plots combine all optimal predictions for the 15 conserved verified targets in all included homologs of AbcR1 (**C**) or target mRNAs (**D**). Distinct interaction domains are indicated by local maxima marked with upright lines. Below the density plots, schematic alignments of the AbcR1 homologs (**C**) and the targets (**D**) are drawn to visualize the predicted optimal and suboptimal interactions for each organism. Each alignment contains eight lines, one for each organism included in the CopraRNA prediction. The order of the organisms in the alignment from top to bottom is: *A. tumefaciens* C58 (*A.t.*), *A. radiobacter* K84 (*A.r.*), *R. etli* CFN42 (*R.e.*), *R. leguminosarum* bv. vicae (*R.l.*), *R. etli* 652 (*R.e.*652) *S. meliloti* 1021 (*S.m.*), and *S. medicae* WSM419 (*S.md.*). The aligned regions are colored in gray and the optimal predicted interaction regions are given in different colors (for contrast only). The respective best suboptimal interaction site predictions are additionally shown by gray lines. White regions indicate gaps inside the AbcR1 alignment. Locus tag and gene name (if available) of target mRNAs are given on the right. A vertical gray line indicates the start codon. Numbering of bases in mRNA alignments is given relative to the start codon (**D**). For detailed visualization of optimal and suboptimal interactions for AbcR1 and its verified target mRNAs in *A. tumefaciens*, see **Table S3**.

responsible for regulation of *atu2422* and *atu3114* (**Fig. 8B and D**). In accordance with the band shift results (**Fig. 6C**), *atu4678* was predominantly controlled via M2 (**Fig. 8C**). The same was true for *atu4431*, which was predicted to bind the M2 region AbcR1 in in its coding sequence (**Fig. 8E**).

Binding sites of AbcR1 in the CDS contain SD-like sequences

To precisely map the AbcR1-binding positions in the CDS of selected target mRNAs, we used an in vitro reverse transcription approach. The principle is illustrated in **Figure 9A**. Target mRNA fragments of 100 or 150 nt length were annealed to end-labeled primers complementary to regions upstream of the predicted interaction region followed by cDNA synthesis. Truncated products upon addition of two different concentrations of AbcR1 prior to reverse transcription were mapped in reference to a sequence reaction run on the same gel. In a control experiment (**Fig. 9B**), the mapped *atu2422*–AbcR1 interaction site corresponded to the previously reported site overlapping the SD sequence of the mRNA.[26] As an example for an mRNA targeted far within the CDS, *atu3114* was used. The CopraRNA-predicted region around 516 nt in the open reading frame was found to interact with AbcR1, thus resulting in prematurely terminated cDNA fragments (**Fig. 9C**). With the *malE* RNA, the presence of AbcR1 led to truncated cDNA products corresponding to a CDS region around +207 (**Fig. 9D**). The mapped interactions sites show that AbcR1 addresses SD-like UGGGAG motifs (see sequence to the left of **Fig. 9B–D**) regardless of their position in the mRNA.

AbcR1 promotes degradation of target mRNAs when bound to the TIR or CDS

The previously identified target mRNAs of *atu2422* and *atu1879* were significantly stabilized in the absence of AbcR1 in vivo.[26] This provided evidence that interaction of M1 with the TIR (*atu2422*) or CDS (*atu1879*) accelerates mRNA turnover and led us to study the effect of AbcR1 on the stability of various target mRNAs. We selected one example each for M1-TIR, M2-TIR, M1-CDS, M2-TIR, and M1/M2-CDS interactions and determined mRNA degradation in the presence or absence of AbcR1 after transcription was stopped by addition of rifampicin.

Interaction of AbcR1 with the TIR via M1 (*frcB*, **Fig. 10A**) or M2 (*atu4678*, **Fig. 10B**) destabilizes the target mRNAs as shown by their elevated stability in the absence of the sRNA. The same is true when the CDS is bound by AbcR1 either by M1 (*atu3114*, **Fig. 10C**), M2 (*atu4431*, **Fig. 10D**), or M1 or M2 (*malE*, **Fig. 10E**) suggesting that negative regulation by AbcR1 involves RNA degradation regardless of whether the TIR or CDS is targeted. Contrary to these negatively regulated transcripts, stability of the positively regulated *atpH* transcripts was not influenced by AbcR1 (**Fig. 10F**).

Discussion

Global approaches like proteomics or microarrays and bioinformatic predictions are commonly used for sRNA target identification.[2,8,67] In this study, we employed a combination of global proteomics and comparative biocomputational predictions for identifying targets of AbcR1 in the plant-pathogen *A. tumefaciens*. Validation of 14 targets via northern blot hybridization enlarged the set of currently known AbcR1 targets to 16 mRNAs. Although several target mRNAs of AbcR1 have been reported in *Brucella* and *Sinorhizobium*, the mode of action of this conserved sRNA has not yet been studied.[57,59] Our study uncovered two distinct target-binding sites in AbcR1 and variable interacting loci in the controlled transcripts.

AbcR1 targets different sites of mRNAs through two functional modules

Many Hfq-associated sRNAs contain one single-stranded domain able to interact with multiple target mRNAs.[5,14,68-71] Other sRNAs have several functional domains that base pair with different sets of target mRNAs in *E. coli*, *Salmonella*, and *Vibrio harveyi*.[7,16,72,73]

Previously, only one conserved target-binding region strategically positioned in the first exposed hairpin loop of AbcR1 has been reported.[26,59] Here, we exploited the recently established comparative target prediction tool CopraRNA, which has been previously used to predict the two and three interaction regions of GcvB and Spot42, respectively.[7,8,73] Strikingly, the two computationally predicted and experimentally verified modules M1 and M2 are highly conserved among *Rhizobiaceae* suggesting that the two functional modules are not limited to AbcR1 in *A. tumefaciens*. The more distantly related AbcR1 sequence from *Brucella* was not included in the CopraRNA predictions because it exhibits less sequence identity to *A. tumefaciens* AbcR1 than

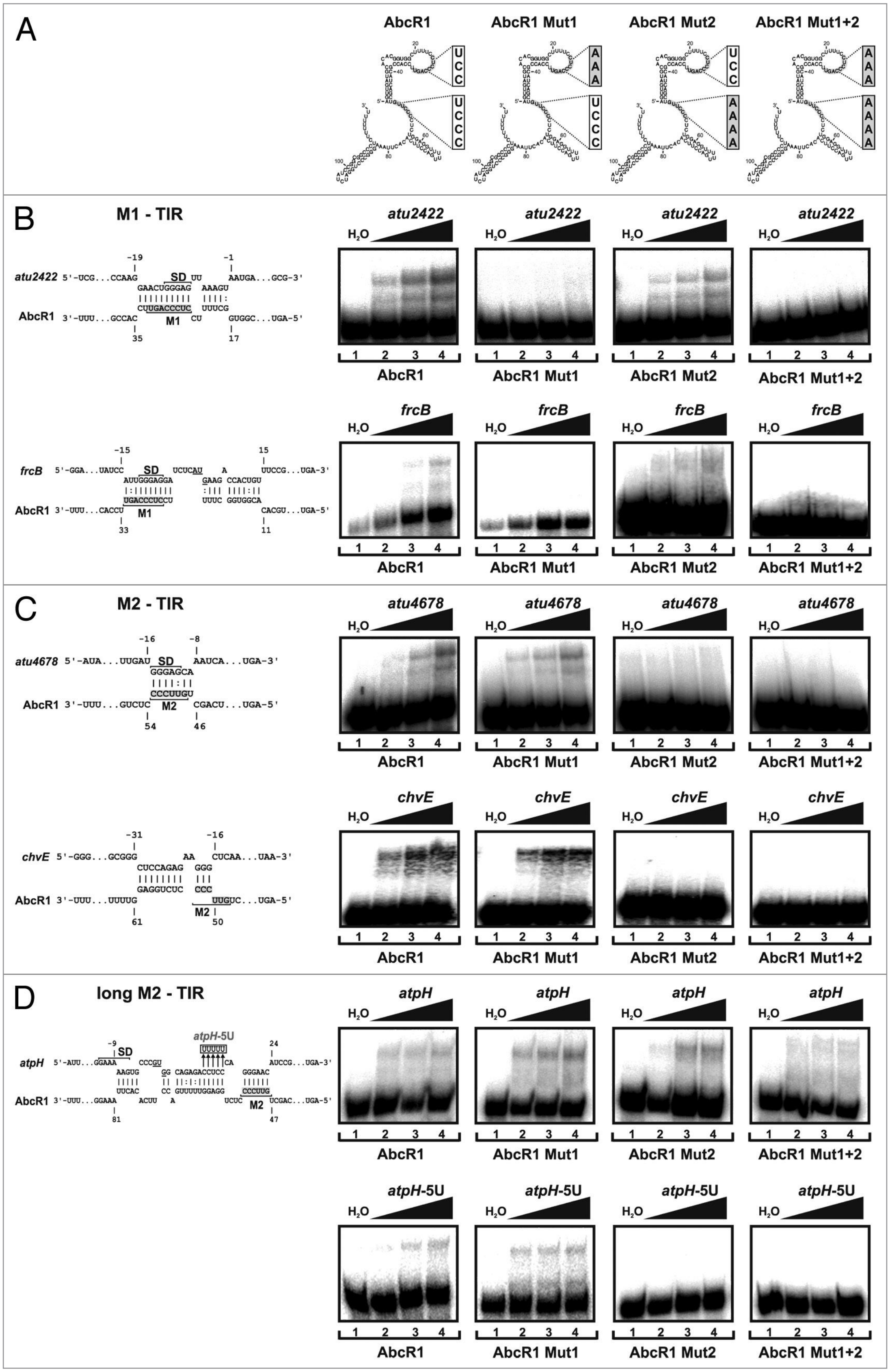

A
AbcR1
AbcR1 Mut1
AbcR1 Mut2
AbcR1 Mut1+2
U C C
A A A
B
M1 - TIR
-19
-1
atu2422 5'-UCG...CCAAG SD UU AAUGA...GCG-3'
GAACUGGGAG AAAGU
CUUGACCCUC UUUCG
AbcR1 3'-UUU...GCCAC M1 CU GUGGC...UGA-5'
35
17
-15
15
frcB 5'-GGA...UAUCC SD UCUCAU A UUCCG...UGA-3'
AUUGGGAGGA GAAG CCACUGU
UGACCCUCCU UUUC GGUGGCA
AbcR1 3'-UUU...CACCU M1 CACGU...UGA-5'
33
11
atu2422
frcB
H2O
1 2 3 4
AbcR1
AbcR1 Mut1
AbcR1 Mut2
AbcR1 Mut1+2
C
M2 - TIR
-16
-8
atu4678 5'-AUA...UUGAU SD AAUCA...UGA-3'
GGGAGCA
CCCUUGU
AbcR1 3'-UUU...GUCUC M2 CGACU...UGA-5'
54
46
-31
-16
chvE 5'-GGG...GCGGG AA CUCAA...UAA-3'
CUCCAGAG GGG
GAGGUCUC CCC
AbcR1 3'-UUU...UUUUG UUGUC...UGA-5'
M2
61
50
atu4678
chvE
D
long M2 - TIR
atpH-5U
UUUUU
-9
SD
24
atpH 5'-AUU...GGAAA CCCGU CA AUCCG...UGA-3'
AAGUG GG CAGAGACCUCC GGGAAC
UUCAC CC GUUUUUGGAGG CCCUUG
AbcR1 3'-UUU...GGAAA ACUU A UCUC M2 UCGAC...UGA-5'
81
47
atpH
atpH-5U

Figure 6. For figure legend, see page 633.

Figure 6 (See opposite page). Binding of target mRNAs at the translation initiation region by two distinct functional modules. (**A**) Secondary structures of WT AbcR1 and the variants Mut1, Mut2, and Mut1+2. Band shift experiments with AbcR1 variants and *atu2422* (**B**), *frcB* (**B**), *atu4678* (**C**), *chvE* (**C**), and *atpH* (**D**) mRNA fragments (~-50/+100 nt relative to the AUG start codon). Predicted IntaRNA duplexes formed by AbcR1 and target mRNAs are shown to the left. Numbering of mRNA nucleotides is given relative to the AUG/GUG start codon. ^{32}P-labeled AbcR1 variants (< 0.05 pmol) were incubated with increasing concentrations of unlabeled target RNAs at 30 °C for 20 min. Final concentrations of unlabeled RNA were added in 100 (lanes 2), 200 (lanes 3), and 400 (lanes 4) fold excess. Samples shown in lanes 1 were incubated with water (control).

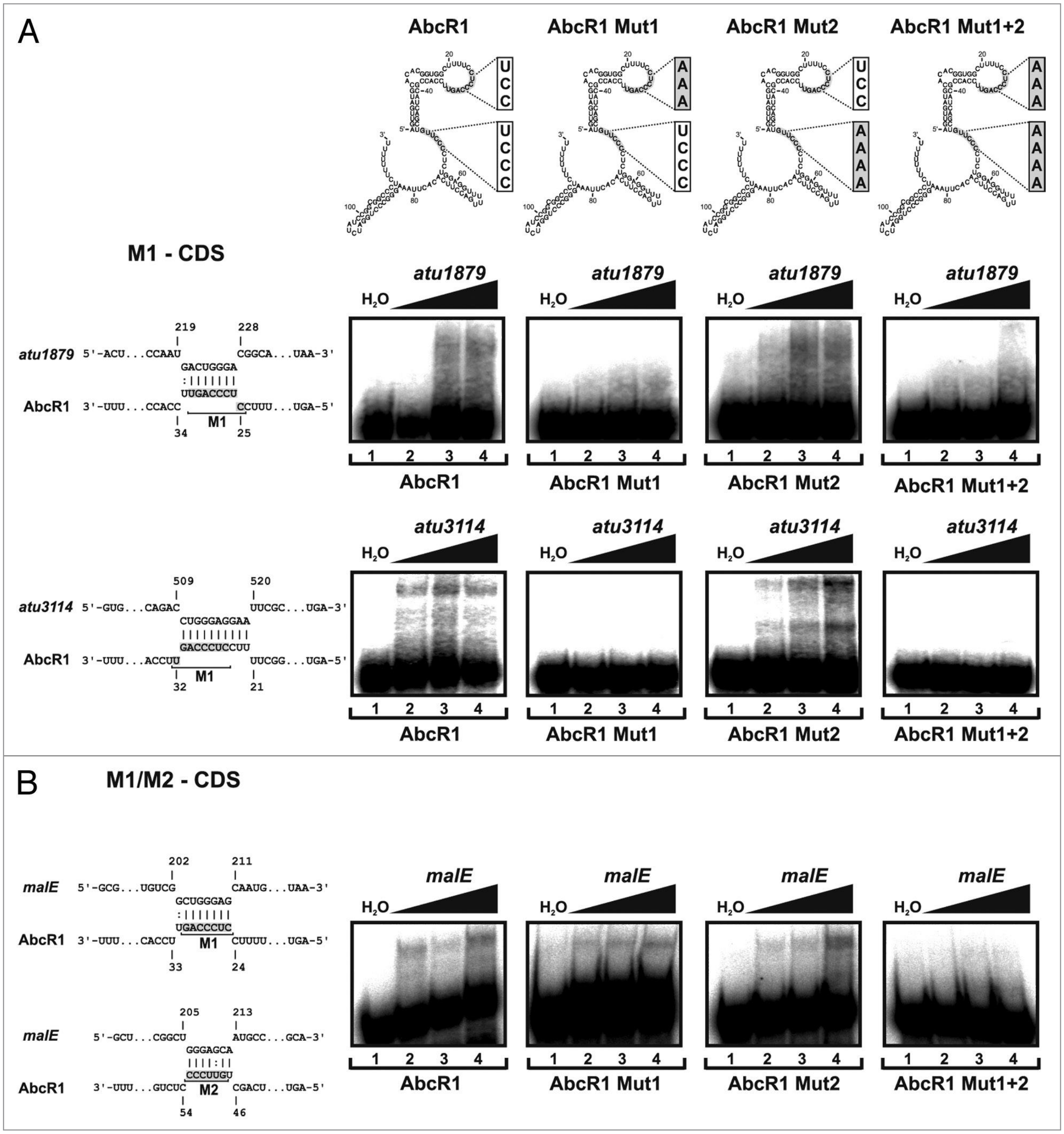

A
AbcR1
AbcR1 Mut1
AbcR1 Mut2
AbcR1 Mut1+2
U C C
U C C C
A A A
A A A A
20
40
60
80
100
5'-AUG
3'
M1 - CDS
219
228
atu1879 5'-ACU...CCAAU CGGCA...UAA-3'
GACUGGGA
:|||||||
UUGACCCU
AbcR1 3'-UUU...CCACC CCUUU...UGA-5'
M1
34
25
atu1879
H2O
1 2 3 4
509
520
atu3114 5'-GUG...CAGAC UUCGC...UGA-3'
CUGGGAGGAA
||||||||||
GACCCUCCUU
AbcR1 3'-UUU...ACCUU UUCGG...UGA-5'
M1
32
21
atu3114
B
M1/M2 - CDS
202
211
malE 5'-GCG...UGUCG CAAUG...UAA-3'
GCUGGGAG
:|||||||
UGACCCUC
AbcR1 3'-UUU...CACCU CUUUU...UGA-5'
M1
33
24
205
213
malE 5'-GCU...CGGCU AUGCC...GCA-3'
GGGAGCA
||||:||
CCCUUGU
AbcR1 3'-UUU...GUCUC CGACU...UGA-5'
M2
54
46
malE

Figure 7. Binding of target mRNAs in the coding sequence by AbcR1. (Top) Secondary structures of AbcR1 wild-type, the variants Mut1, Mut2, and Mut1+2. Band shift experiments with AbcR1 variants and *atu1879* (**A**), *atu3114* (**A**), and *malE* (**B**) mRNA fragments (~150 nt). Predicted IntaRNA duplexes formed by AbcR1 and target mRNAs are shown to the left. Numbering of mRNA nucleotides is given relative to the start codon. ^{32}P-labeled AbcR1 variants (< 0.05 pmol) were incubated with increasing concentrations of unlabeled target RNAs at 30 °C for 20 min. Final concentrations of unlabeled RNA were added in 100 (lanes 2), 200 (lanes 3) and 400 (lanes 4) fold excess. Samples from lanes 1 were incubated with water (control).

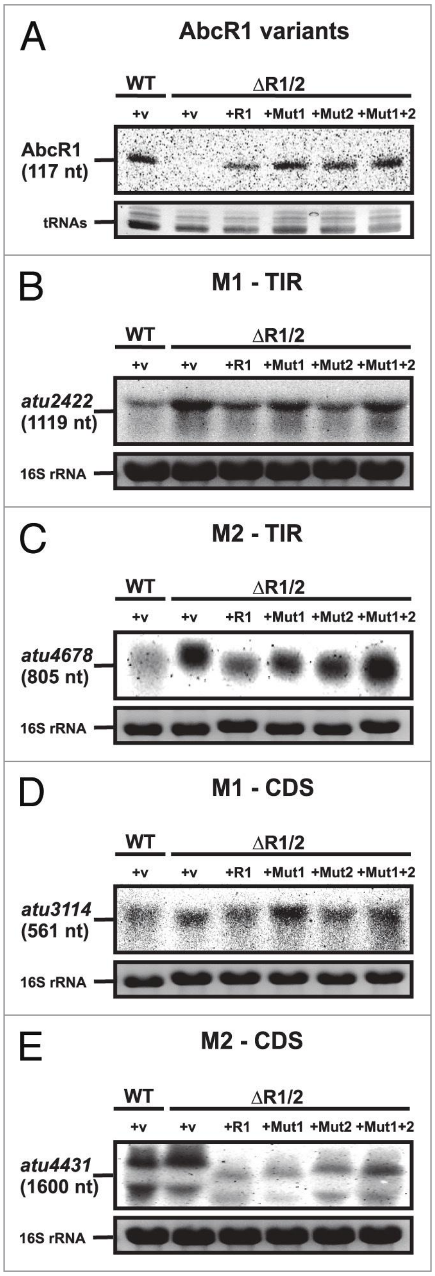


Figure 8. In vivo validation of AbcR1 modules 1 and 2. Northern blot analyses of AbcR1 (**A**), *atu2422* (**B**), *atu4678* (**C**), *atu3114* (**D**), and *atu4431* (**E**) transcripts from cultures of the *A. tumefaciens* wild-type (WT) or the ΔAbcR1/2 deletion mutant (ΔR1/2) complemented with a plasmid expressing different AbcR1 variants (+R1, +Mut1, +Mut2, +Mut1+2). The strains were grown in YEB medium. +v: control strains harboring the empty vector. Eight μg of total RNA were separated on 1.2% denaturing agarose gels. Ethidiumbromide-stained tRNAs or 16S rRNAs were used as loading control.

the homologs from *Sinorhizobium* and *Rhizobium* species. The existence of two single-stranded M1- and M2-like regions in the predicted secondary structure of *B. abortus* AbcR1, however, suggests that two functional AbcR1 modules are not restricted to plant-associated bacteria.[59]

On the target site, our study revealed that AbcR1 binding regions are scattered throughout the TIR and CDS. Although interference with translation by mRNA binding around the SD sequence is considered the most common control mechanism of sRNAs, targeting of coding sequences has been described in enterobacteria, for example, MicC and *ompD*, ArcZ-*tpx*, RybB-*fadL*, SgrS-*manX*, and SdsR-*ompD*.[14,71,74-76] Two exposed UC-rich interaction regions in AbcR1 and the potential to interact with SD-like regions in the TIR or CDS allows pervasive gene regulation by this sRNA in *A. tumefaciens*.

AbcR1: A conserved master regulator of ABC transporters

There is increasing evidence that sRNAs are more than single target regulators, but rather act on multiple *trans*-encoded targets and rewire entire transcriptional networks.[2,77,78] Many well-studied sRNAs in enterobacteria control large sets of functionally related target mRNAs; for example, RyhB regulates mRNAs encoding iron-binding proteins involved in iron homeostasis, OmrA/OmrB regulate mRNAs encoding proteins for outer membrane protein synthesis, and GcvB controls genes for amino acid biosynthesis and transport.[7,22-24,69,70,79-83]

Homologs of AbcR1 from *S. meliloti*, *R. etli*, and *B. abortus* are similar in sequence and structure.[52-54,56,59,84] A functional classification of target mRNAs (ABC transport system) was initially described for the AbcR sRNAs in *A. tumefaciens* and *B. abortus* 2308.[26,59] The experimental verification of 14 AbcR1 targets encoding periplasmic transport proteins carrying sugars, amino acids, and opines supports the function of AbcR1 as a key regulator for these transport systems (**Fig. 11**).

Our previous study described a potential role of AbcR1 in plant defense, quorum sensing, and virulence of *A. tumefaciens* because the AbcR1 target *atu2422* codes for binding protein of an importer of GABA, a plant defense molecule.[26,85-88] We now find that AbcR1 also silences synthesis of ChvE (**Fig. 4C**), a regulator in sugar-dependent activation of the virulence cascade as well as other virulence-related ABC transporters (NocT and AttC).[43,45] This strengthens the hypothesis that AbcR1 is involved in plant–microbe interactions and post-infection nutrient acquisition.

Although most currently known sRNAs block translation of target mRNAs by interfering with ribosome binding, several sRNAs can activate gene expression.[89] They can, for example, bind upstream of the TIR and remodel an intrinsic inhibitory mRNA structure such that the sequestered ribosome binding site is liberated (DsrA and RprA).[90-92] Recently, new translation-independent pathways of mRNA activations have been reported for *cfa* through RydC and for *yigL* through SgrS in *Salmonella*.[19,93] In enterobacteria, well-characterized sRNAs like RyhB and ArcZ repress some target mRNAs, but activate translation of *shiA* (RyhB) and *rpoS* (ArcZ).[69,74,79,80,94] In addition to the many negatively controlled AbcR1 targets in *A. tumefaciens* we found *atpH* as positively regulated gene. It is predicted to encode the delta subunit of the ATP synthase. In *E. coli*, this subunit plays a key role in the assembly of the H^+-translocating F_0F_1 ATP synthase.[95] Although it remains unknown how AbcR1 controls *atpH* expression, for instance, AbcR1 does not alter *atpH* mRNA stability directly, the extensive interaction region between AbcR1 M2 and the TIR of *atpH* is indicative of a direct mechanism. Control of multiple ABC transporters and the ATP synthase suggests that AbcR1 coordinates nutrient acquisition and energy conversion in *A. tumefaciens*.

Experimental Procedures

Bacterial growth conditions

Bacterial strains and antibiotics used in this study are listed in **Table S5**. *E. coli* was grown in LB medium at 37 °C. *A. tumefaciens* strains were cultivated in YEB medium at 30 °C.

Strain and vector constructions

The ΔAbcR1 and Δ*hfq* mutant strains were constructed in previous studies.[26,61] Runoff plasmids as templates for in vitro transcription of AbcR1 or target mRNA fragments were flanked by the T7-promoter sequence (GAAATTAATA CGACTCACTA TAGGG) and an *Eco*RV site PCR-amplified with primers listed in **Table S4** and subcloned into pUC18.[96] AbcR1 variants were constructed via site-directed mutagenesis using the primers listed in **Table S4**.

Protein preparation

Cells of *A. tumefaciens* wild-type, ΔAbcR1, and Δ*hfq* were grown in 30 ml YEB medium at 30 °C to an OD600 of 1.5. Culture volumes of 30 ml were harvested, washed three times in 30 ml of TE-buffer (100 mM Tris and 1 mM EDTA), and finally resuspended in 4 ml of TE-buffer with 1.39 mM PMSF and 0.2 mM DTT. Cells were disrupted by three passes through a chilled French press. The lysates were centrifuged at 10 000 × g for 30 min to remove the cell debris. Protein concentrations were determined by Bradford assays.[97]

Two-dimensional PAGE and Mass Spectrometry

Total proteins extracts of *A. tumefaciens* wild-type, ΔAbcR1, and Δ*hfq* cells were concentrated by chloroform/methanol precipitation up to 600 µg µl-1.[98] Isoelectric focusing and SDS-PAGE were performed as described previously.[99] Protein solutions were loaded on Immobiline DryStrip pH 4–7, 24 cm (GE Healthcare). After isoelectric focusing, proteins were subjected to 12.5% SDS-PAGE, and the spots were visualized using RuBPS ($C_{72}H_{42}N_6Na_4O_{18}RuS_6$) staining. Protein spots were scanned using a Typhoon TRIO (GE Healthcare) and were quantified with the Delta two-dimensional software (version 4.0, Decodon). Selected protein spots were excised from the gel, and protein identification

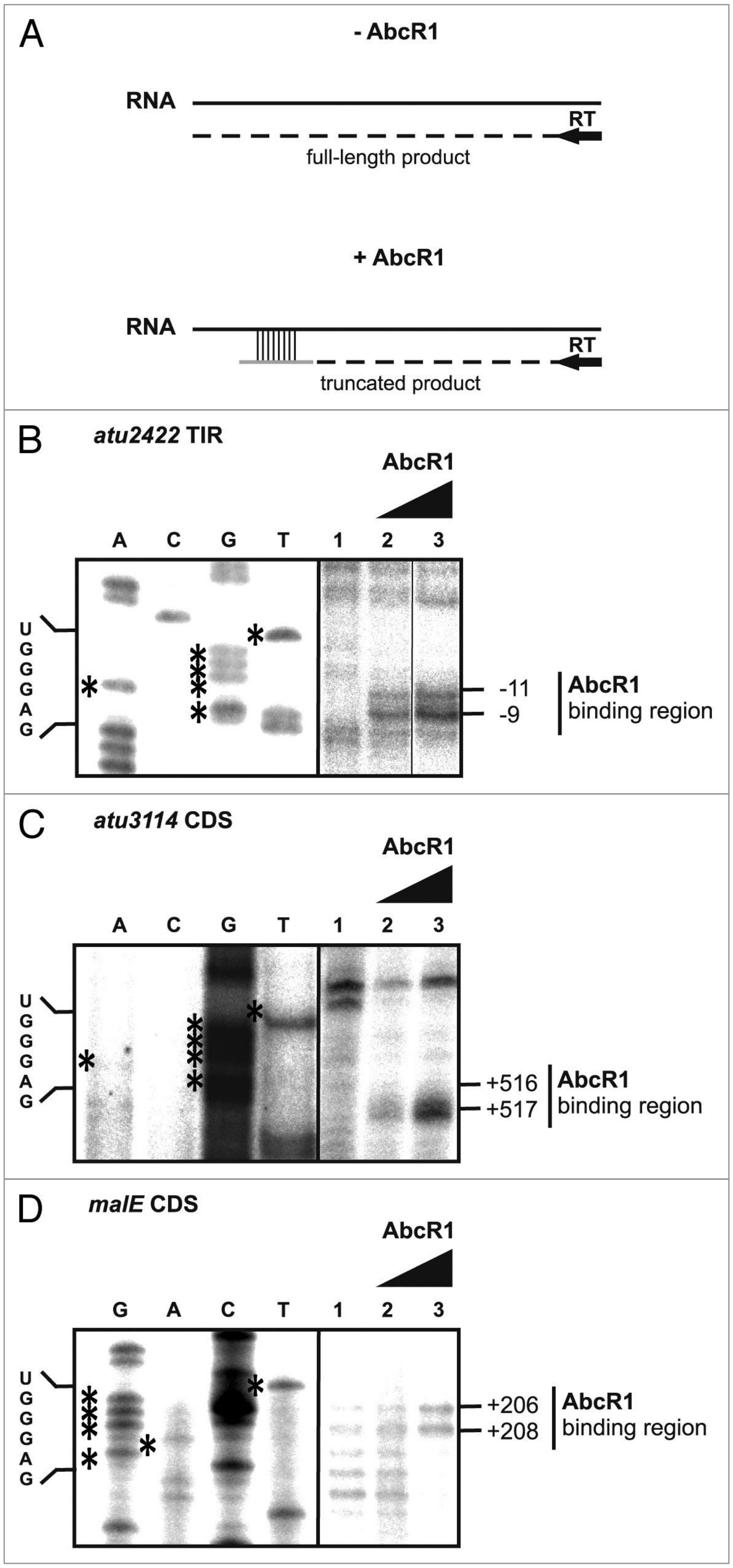


Figure 9. Precise mapping of AbcR1 binding sites in target mRNAs. (**A**) Principle of AbcR1 binding-site mapping by toeprinting analysis. -AbcR1, without AbcR1; reverse transcription (RT) starting from a primer complementary to the target mRNA sequence transcribes a full-length product. +AbcR1, pairing of AbcR1 with the target sequence terminates reverse transcription (truncated product). AbcR1 binding-site mapping on *atu2422* (**B**), *atu3114* (**C**), and *malE* (**D**) RNA fragments was performed as described in *Experimental procedures*. The position of truncated products is indicated to the right. mRNA nucleotides involved in M1 binding are shown to the left. Concentrations of AbcR1 RNAs were 1.5 pmol µl^{-1} (lane 2) and 2.5 pmol µl^{-1} (lane 3).

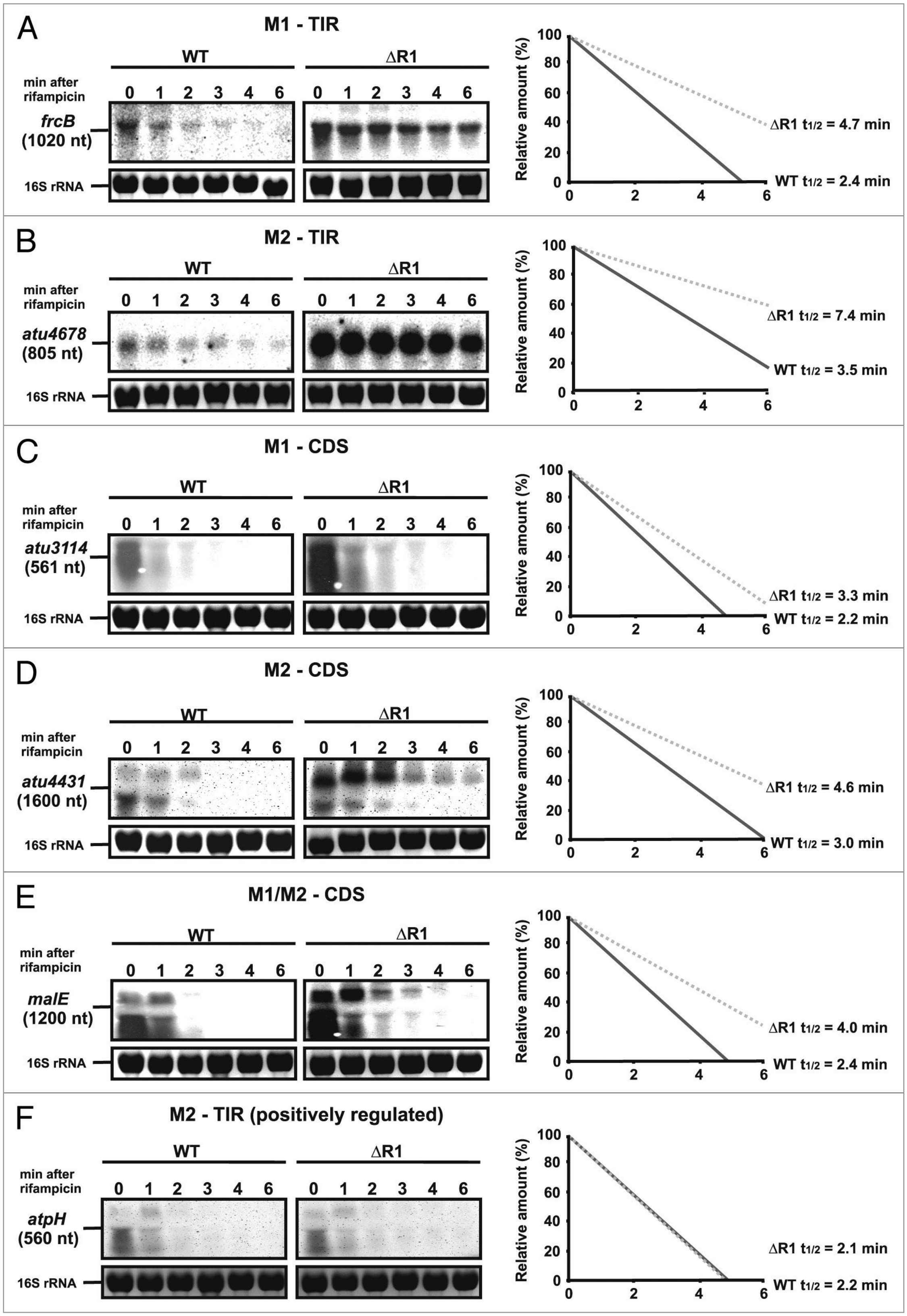


Figure 10. AbcR1 M1 and M2 target mRNAs in the TIR and the CDS for degradation. Northern blot analyses of *frcB* (**A**), *atu4678* (**B**), *atu3114* (**C**), *atu4431* (**D**), *malE* (**E**), and *atpH* (**F**) transcripts from cultures treated with rifampicin. Cultures of the *A. tumefaciens* wild-type (WT) or the ΔAbcR1 deletion mutant (ΔR1) were grown to exponential or stationary (in case of *frcB*) growth phase in YEB medium and treated with rifampicin (250 mg ml^{-1}). Total RNA fractions were collected at the indicated time points. Eight μg of total RNA were separated on 1.2% denaturing agarose gels. Ethidiumbromide-stained 16S rRNAs were used as loading control. Quantification of transcript stabilities and their calculated half-lives are given to the right.

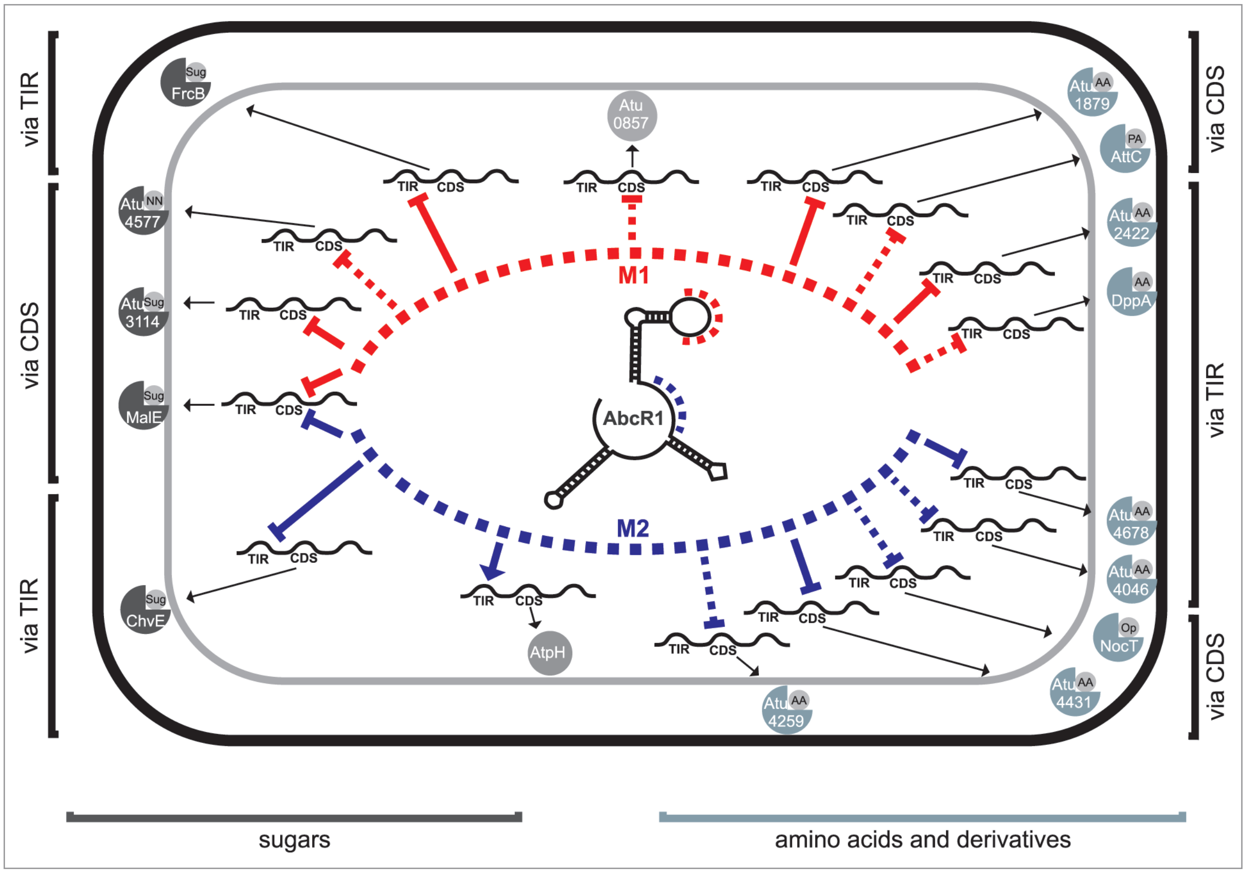


Figure 11. The AbcR1 regulon of *A. tumefaciens*. AbcR1 controls mRNAs of periplasmic substrate-binding proteins of 14 ABC transporters (sugars and amino acids to the left and right, respectively), an annotated oxidoreductase (*atu0857*) and AtpH. Module 1 (red) and module 2 (blue) dependent genes are sorted toward the top and bottom of the schematic cell, respectively. The interaction region (TIR or CDS) and the mode of action (repression or activation) are indicated. Dashed lines refer to computationally predicted interactions.

using mass spectrometry was performed by MALDI-TOF mass spectrometry as described previously.[100]

RNA preparation and northern analysis

Cells were harvested, washed, and frozen in liquid nitrogen as described previously.[26] Isolation of total RNA was done by using the hot acid phenol method.[101] Northern analyses were performed as previously described.[26] To measure mRNA stability, rifampicin was added to the cell cultures in a final concentration of 250 mg ml^{-1} and samples for RNA isolation were collected before (0 min) and 1, 2, 3, 4, and 6 min after addition of the transcriptional inhibitor rifampicin. In order to determine the half-life of the specific mRNAs, the amount of transcripts present at each time point was quantified using the Image software Alpha Ease FC (Alpha Innotech). The primers used for RNA probe generation are listed in **Table S4** in the supplemental material.

Gel shift experiments

The sRNAs AbcR1 WT, QC1, QC2, and QC1+2 and the target mRNA fragments (comprising ~150 nucleotides in the TIR or in the CDS) were synthesized in vitro by runoff transcription with T7 RNA polymerase from the linearized plasmids listed in **Table S4**. 5‘ end labeling of AbcR1 WT or AbcR1 variants (QC1, QC2, and QC1+2) with ^{32}P was performed as described.[102] RNA band shift experiments were performed in 1x structure buffer (Ambion) in a total reaction mixture volume of 15 µl as follows. 5′ end labeled AbcR1 (corresponding to 5000 c.p.m.) and 1 µg of tRNA were incubated in the presence of unlabeled target mRNA fragments (~150 nt) at 30 °C for 20 min. The final concentrations of added unlabeled RNA fragments are given in the figure legends. Prior to gel loading, the binding reactions were mixed 4.5 µl of native loading dye (50% glycerol, 0.5× TBE, 0.1% bromophenol blue and 0.1% xylene cyanol) and run on native 6% polyacrylamide gels in 0.5× TBE buffer at 300 V for 1.5–3 h.

Mapping of sRNA-binding sites

Mapping of AbcR1-binding sites were performed like previously described "toeprint analysis" with some modifications.[103] Annealing mixtures contained 0.5 pmol unlabeled *atu2422* (50 nt +/– from the AUG start codon), *malE* (+100/+250 relative to AUG start codon), or *atu3114* (+437/+588 relative to AUG start codon) mRNA fragments and 1 pmol of 5′ end labeled primer runoff_*atu2422*_rv, runoff_*malE2*_rv, and runoff_*atu3114*_rv in VD buffer without magnesium. Annealing mixtures were heated for 3 min at 80 °C and snap frozen in a frozen plastic box. After incubation on ice for 20 min, different concentrations (listed in figure legends) of AbcR1, WT, or water (as negative control) were added and incubated at 37 °C for 20 min. After addition of 2 µl MMLV-Mix (VD + Mg^{2+}, BSA, dNTPs and MMLV reverse transcriptase [USB]), cDNA synthesis were performed at 37 °C for 10 min. Reactions were stopped by adding formamide loading dye and reaction aliquots were separated on a denaturing 8% polyacrylamide gel. Reverse transcription cDNA products were identified by comparison with sequences generated with the same 5′ end labeled primer.

Bioinformatic tools

Alignments of sequences were generated by the ClustalW software obtained from http://www.ebi.ac.uk/Tools/msa/clustalw2/. sRNA–mRNA duplexes were predicted by the IntaRNA webserver from http://rna.informatik.uni-freiburg.de:8080/v1/IntaRNA.jsp.[104] Secondary structures and consensus structures were computed with mfold http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form and RNAalifold http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi.[65,105] Comparative predictions for AbcR1 binding sites and targets were done with CopraRNA http://rna.informatik.uni-freiburg.de/CopraRNA/Input.jsp.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Rosemarie Gurski for excellent technical assistance, Nicole Frankenberg-Dinkel and Julia Bandow for sharing proteomic infrastructure and expertise; Knut Büttner (Greifswald, Germany) for MALDI-MS analysis and Philip Möller for critical reading of the manuscript. Parts of this study were conducted in the Central Unit for Ion beams and Radionucleotides (RUBION). This work was supported by a grant from the German Research Foundation (DFG priority program SPP 1258: Sensory and regulatory RNAs in Prokaryotes) to F.N. and W.R.H. and by the Bundesministerium für Bildung und Forschung (BMBF) Grant 0316165 to W.R.H.

Supplemental Material

Supplemental material may be found here:
www.landesbioscience.com/journals/rnabiology/article/29145/

References

1. Lalaouna D, Simoneau-Roy M, Lafontaine D, Massé E. Regulatory RNAs and target mRNA decay in prokaryotes. Biochim Biophys Acta 2013; 1829:742-7; PMID:23500183; http://dx.doi.org/10.1016/j.bbagrm.2013.02.013
2. Storz G, Vogel J, Wassarman KM. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 2011; 43:880-91; PMID:21925377; http://dx.doi.org/10.1016/j.molcel.2011.08.022
3. Kawamoto H, Koide Y, Morita T, Aiba H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. Mol Microbiol 2006; 61:1013-22; PMID:16859494; http://dx.doi.org/10.1111/j.1365-2958.2006.05288.x
4. Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J. Small RNA binding to 5′ mRNA coding region inhibits translational initiation. Mol Cell 2008; 32:827-37; PMID:19111662; http://dx.doi.org/10.1016/j.molcel.2008.10.027
5. Balbontín R, Fiorini F, Figueroa-Bossi N, Casadesús J, Bossi L. Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in *Salmonella enterica.* Mol Microbiol 2010; 78:380-94; PMID:20979336; http://dx.doi.org/10.1111/j.1365-2958.2010.07342.x
6. Boehm A, Vogel J. The *csgD* mRNA as a hub for signal integration via multiple small RNAs. Mol Microbiol 2012; 84:1-5; PMID:22414234; http://dx.doi.org/10.1111/j.1365-2958.2012.08033.x
7. Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JCD, Vogel J. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. Mol Microbiol 2011; 81:1144-65; PMID:21696468; http://dx.doi.org/10.1111/j.1365-2958.2011.07751.x
8. Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J. Comparative genomics boosts target prediction for bacterial small RNAs. Proc Natl Acad Sci U S A 2013; 110:E3487-96; PMID:23980183; http://dx.doi.org/10.1073/pnas.1303248110
9. Bouché F, Bouché JP. Genetic evidence that DicF, a second division inhibitor encoded by the *Escherichia coli dicB* operon, is probably RNA. Mol Microbiol 1989; 3:991-4; PMID:2477663; http://dx.doi.org/10.1111/j.1365-2958.1989.tb00249.x
10. Wassarman KM, Storz G. 6S RNA regulates *E. coli* RNA polymerase activity. Cell 2000; 101:613-23; PMID:10892648; http://dx.doi.org/10.1016/S0092-8674(00)80873-9
11. Mank NN, Berghoff BA, Hermanns YN, Klug G. Regulation of bacterial photosynthesis genes by the small noncoding RNA PcrZ. Proc Natl Acad Sci U S A 2012; 109:16306-11; PMID:22988125; http://dx.doi.org/10.1073/pnas.1207067109
12. Altuvia S, Zhang A, Argaman L, Tiwari A, Storz G. The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. EMBO J 1998; 17:6069-75; PMID:9774350; http://dx.doi.org/10.1093/emboj/17.20.6069
13. Harris JF, Micheva-Viteva S, Li N, Hong-Geller E. Small RNA-mediated regulation of host-pathogen interactions. Virulence 2013; 4:785-95; PMID:23958954; http://dx.doi.org/10.4161/viru.26119
14. Papenfort K, Bouvier M, Mika F, Sharma CM, Vogel J. Evidence for an autonomous 5′ target recognition domain in an Hfq-associated small RNA. Proc Natl Acad Sci U S A 2010; 107:20435-40; PMID:21059903; http://dx.doi.org/10.1073/pnas.1009784107
15. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae.* Cell 2004; 118:69-82; PMID:15242645; http://dx.doi.org/10.1016/j.cell.2004.06.009
16. Shao Y, Feng L, Rutherford ST, Papenfort K, Bassler BL. Functional determinants of the quorum-sensing non-coding RNAs and their roles in target regulation. EMBO J 2013; 32:2158-71; PMID:23838640; http://dx.doi.org/10.1038/emboj.2013.155
17. Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, Baker CS, Georgellis D, Babitzke P, Romeo T. A novel sRNA component of the carbon storage regulatory system of *Escherichia coli.* Mol Microbiol 2003; 48:657-70; PMID:12694612; http://dx.doi.org/10.1046/j.1365-2958.2003.03459.x
18. Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, Camacho MI, Fields JA, Thompson SA, Georgellis D, et al. Circuitry linking the Csr and stringent response global regulatory systems. Mol Microbiol 2011; 80:1561-80; PMID:21488981; http://dx.doi.org/10.1111/j.1365-2958.2011.07663.x
19. Papenfort K, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J. Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. Cell 2013; 153:426-37; PMID:23582330; http://dx.doi.org/10.1016/j.cell.2013.03.003
20. Göpel Y, Papenfort K, Reichenbach B, Vogel J, Görke B. Targeted decay of a regulatory small RNA by an adaptor protein for RNase E and counteraction by an anti-adaptor RNA. Genes Dev 2013; 27:552-64; PMID:23475961; http://dx.doi.org/10.1101/gad.210112.112
21. Sun Y, Vanderpool CK. Physiological consequences of multiple-target regulation by the small RNA SgrS in *Escherichia coli.* J Bacteriol 2013; 195:4804-15; PMID:23873911; http://dx.doi.org/10.1128/JB.00722-13
22. Urbanowski ML, Stauffer LT, Stauffer GV. The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli.* Mol Microbiol 2000; 37:856-68; PMID:10972807; http://dx.doi.org/10.1046/j.1365-2958.2000.02051.x
23. Sharma CM, Darfeuille F, Plantinga TH, Vogel J. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Genes Dev 2007; 21:2804-17; PMID:17974919; http://dx.doi.org/10.1101/gad.447207
24. Pulvermacher SC, Stauffer LT, Stauffer GV. Role of the sRNA GcvB in regulation of *cycA* in *Escherichia coli.* Microbiology 2009; 155:106-14; PMID:19118351; http://dx.doi.org/10.1099/mic.0.023598-0
25. Antal M, Bordeau V, Douchin V, Felden B. A small bacterial RNA regulates a putative ABC transporter. J Biol Chem 2005; 280:7901-8; PMID:15618228; http://dx.doi.org/10.1074/jbc.M413071200
26. Wilms I, Voss B, Hess WR, Leichert LI, Narberhaus F. Small RNA-mediated control of the *Agrobacterium tumefaciens* GABA binding protein. Mol Microbiol 2011; 80:492-506; PMID:21320185; http://dx.doi.org/10.1111/j.1365-2958.2011.07589.x
27. Saier MH Jr. Families of transmembrane sugar transport proteins. Mol Microbiol 2000; 35:699-710; PMID:10692148; http://dx.doi.org/10.1046/j.1365-2958.2000.01759.x
28. Davidson AL, Chen J. ATP-binding cassette transporters in bacteria. Annu Rev Biochem 2004; 73:241-68; PMID:15189142; http://dx.doi.org/10.1146/annurev.biochem.73.011303.073626

29. Hosie AH, Poole PS. Bacterial ABC transporters of amino acids. Res Microbiol 2001; 152:259-70; PMID:11421273; http://dx.doi.org/10.1016/S0923-2508(01)01197-4

30. Narberhaus F, Vogel J. Regulatory RNAs in prokaryotes: here, there and everywhere. Mol Microbiol 2009; 74:261-9; PMID:19732342; http://dx.doi.org/10.1111/j.1365-2958.2009.06869.x

31. Gimpel M, Heidrich N, Mäder U, Krügel H, Brantl S. A dual-function sRNA from *B. subtilis*: SR1 acts as a peptide encoding mRNA on the *gapA* operon. Mol Microbiol 2010; 76:990-1009; PMID:20444087; http://dx.doi.org/10.1111/j.1365-2958.2010.07158.x

32. Romby P, Charpentier E. An overview of RNAs with regulatory functions in gram-positive bacteria. Cell Mol Life Sci 2010; 67:217-37; PMID:19859665; http://dx.doi.org/10.1007/s00018-009-0162-8

33. Dühring U, Axmann IM, Hess WR, Wilde A. An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. Proc Natl Acad Sci U S A 2006; 103:7054-8; PMID:16636284; http://dx.doi.org/10.1073/pnas.0600927103

34. Jäger D, Pernitzsch SR, Richter AS, Backofen R, Sharma CM, Schmitz RA. An archaeal sRNA targeting *cis*- and *trans*-encoded mRNAs via two distinct domains. Nucleic Acids Res 2012; 40:10964-79; PMID:22965121; http://dx.doi.org/10.1093/nar/gks847

35. Schmidtke C, Abendroth U, Brock J, Serrania J, Becker A, Bonas U. Small RNA sX13: a multifaceted regulator of virulence in the plant pathogen *Xanthomonas*. PLoS Pathog 2013; 9:e1003626; PMID:24068933; http://dx.doi.org/10.1371/journal.ppat.1003626

36. Lee K, Huang X, Yang C, Lee D, Ho V, Nobuta K, Fan JB, Wang K. A genome-wide survey of highly expressed non-coding RNAs and biological validation of selected candidates in *Agrobacterium tumefaciens*. PLoS One 2013; 8:e70720; PMID:23950988; http://dx.doi.org/10.1371/journal.pone.0070720

37. Wilms I, Overlöper A, Nowrousian M, Sharma CM, Narberhaus F. Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen *Agrobacterium tumefaciens*. RNA Biol 2012; 9:446-57; PMID:22336765; http://dx.doi.org/10.4161/rna.17212

38. Pitzschke A, Hirt H. New insights into an old story: *Agrobacterium*-induced tumour formation in plants by plant transformation. EMBO J 2010; 29:1021-32; PMID:20150897; http://dx.doi.org/10.1038/emboj.2010.8

39. Lacroix B, Citovsky V. The roles of bacterial and host plant factors in *Agrobacterium*-mediated genetic transformation. Int J Dev Biol 2013; 57:467-81; PMID:24166430; http://dx.doi.org/10.1387/ijdb.130199bl

40. Zupan J, Muth TR, Draper O, Zambryski P. The transfer of DNA from *agrobacterium tumefaciens* into plants: a feast of fundamental insights. Plant J 2000; 23:11-28; PMID:10929098; http://dx.doi.org/10.1046/j.1365-313x.2000.00808.x

41. McCullen CA, Binns AN. *Agrobacterium tumefaciens* and plant cell interactions and activities required for interkingdom macromolecular transfer. Annu Rev Cell Dev Biol 2006; 22:101-27; PMID:16709150; http://dx.doi.org/10.1146/annurev.cellbio.22.011105.102022

42. Doty SL, Yu MC, Lundin JI, Heath JD, Nester EW. Mutational analysis of the input domain of the VirA protein of *Agrobacterium tumefaciens*. J Bacteriol 1996; 178:961-70; PMID:8576069

43. Hu X, Zhao J, DeGrado WF, Binns AN. *Agrobacterium tumefaciens* recognizes its host environment using ChvE to bind diverse plant sugars as virulence signals. Proc Natl Acad Sci U S A 2013; 110:678-83; PMID:23267119; http://dx.doi.org/10.1073/pnas.1215033110

44. Citovsky V, Kozlovsky SV, Lacroix B, Zaltsman A, Dafny-Yelin M, Vyas S, Tovkach A, Tzfira T. Biological systems of the host cell involved in *Agrobacterium* infection. Cell Microbiol 2007; 9:9-20; PMID:17222189; http://dx.doi.org/10.1111/j.1462-5822.2006.00830.x

45. He F, Nair GR, Soto CS, Chang Y, Hsu L, Ronzone E, DeGrado WF, Binns AN. Molecular basis of ChvE function in sugar binding, sugar utilization, and virulence in *Agrobacterium tumefaciens*. J Bacteriol 2009; 191:5802-13; PMID:19633083; http://dx.doi.org/10.1128/JB.00451-09

46. Matthysse AG, Yarnall HA, Young N. Requirement for genes with homology to ABC transport systems for attachment and virulence of *Agrobacterium tumefaciens*. J Bacteriol 1996; 178:5302-8; PMID:8752352

47. Planamente S, Moréra S, Faure D. In planta fitness-cost of the Atu4232-regulon encoding for a selective GABA-binding sensor in *Agrobacterium*. Commun Integr Biol 2013; 6:e23692; PMID:23710277; http://dx.doi.org/10.4161/cib.23692

48. Kim H, Farrand SK. Characterization of the *acc* operon from the nopaline-type Ti plasmid pTiC58, which encodes utilization of agrocinopines A and B and susceptibility to agrocin 84. J Bacteriol 1997; 179:7559-72; PMID:9393724

49. Hayman GT, Beck von Bodman S, Kim H, Jiang P, Farrand SK. Genetic analysis of the agrocinopine catabolic region of *Agrobacterium tumefaciens* Ti plasmid pTiC58, which encodes genes required for opine and agrocin 84 transport. J Bacteriol 1993; 175:5575-84; PMID:8366042

50. Schneider E, Eckey V, Weidlich D, Wiesemann N, Vahedi-Faridi A, Thaben P, Saenger W. Receptor-transporter interactions of canonical ATP-binding cassette import systems in prokaryotes. Eur J Cell Biol 2012; 91:311-7; PMID:21561685; http://dx.doi.org/10.1016/j.ejcb.2011.02.008

51. Chevrot R, Rosen R, Haudecoeur E, Cirou A, Shelp BJ, Ron E, Faure D. GABA controls the level of quorum-sensing signal in *Agrobacterium tumefaciens*. Proc Natl Acad Sci U S A 2006; 103:7460-4; PMID:16645034; http://dx.doi.org/10.1073/pnas.0600313103

52. del Val C, Rivas E, Torres-Quesada O, Toro N, Jiménez-Zurdo JI. Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. Mol Microbiol 2007; 66:1080-91; PMID:17971083; http://dx.doi.org/10.1111/j.1365-2958.2007.05978.x

53. Ulvé VM, Sevin EW, Chéron A, Barloy-Hubler F. Identification of chromosomal alpha-proteobacterial small RNAs by comparative genome analysis and detection in *Sinorhizobium meliloti* strain 1021. BMC Genomics 2007; 8:467; PMID:18093320; http://dx.doi.org/10.1186/1471-2164-8-467

54. Valverde C, Livny J, Schlüter JP, Reinkensmeier J, Becker A, Parisi G. Prediction of *Sinorhizobium meliloti* sRNA genes and experimental detection in strain 2011. BMC Genomics 2008; 9:416; PMID:18793445; http://dx.doi.org/10.1186/1471-2164-9-416

55. Voss B, Hölscher M, Baumgarth B, Kalbfleisch A, Kaya C, Hess WR, Becker A, Evguenieva-Hackenberg E. Expression of small RNAs in Rhizobiales and protection of a small RNA and its degradation products by Hfq in *Sinorhizobium meliloti*. Biochem Biophys Res Commun 2009; 390:331-6; PMID:19800865; http://dx.doi.org/10.1016/j.bbrc.2009.09.125

56. Vercruysse M, Fauvart M, Cloots L, Engelen K, Thijs IM, Marchal K, Michiels J. Genome-wide detection of predicted non-coding RNAs in *Rhizobium etli* expressed during free-living and host-associated growth using a high-resolution tiling array. BMC Genomics 2010; 11:53; PMID:20089193; http://dx.doi.org/10.1186/1471-2164-11-53

57. Torres-Quesada O, Millán V, Nisa-Martínez R, Bardou F, Crespi M, Toro N, Jiménez-Zurdo JI. Independent activity of the homologous small regulatory RNAs AbcR1 and AbcR2 in the legume symbiont *Sinorhizobium meliloti*. PLoS One 2013; 8:e68147; PMID:23869210; http://dx.doi.org/10.1371/journal.pone.0068147

58. Torres-Quesada O, Reinkensmeier J, Schlüter JP, Robledo M, Peregrina A, Giegerich R, Toro N, Becker A, Jiménez-Zurdo JI. Genome-wide profiling of Hfq-binding RNAs uncovers extensive post-transcriptional rewiring of major stress response and symbiotic regulons in Sinorhizobium meliloti. RNA Biol 2014; 11; PMID:24786641; http://dx.doi.org/10.4161/rna.28239

59. Caswell CC, Gaines JM, Ciborowski P, Smith D, Borchers CH, Roux CM, Sayood K, Dunman PM, Roop Ii RM. Identification of two small regulatory RNAs linked to virulence in *Brucella abortus* 2308. Mol Microbiol 2012; 85:345-60; PMID:22690807; http://dx.doi.org/10.1111/j.1365-2958.2012.08117.x

60. Torres-Quesada O, Oruezabal RI, Peregrina A, Jofré E, Lloret J, Rivilla R, Toro N, Jiménez-Zurdo JI. The *Sinorhizobium meliloti* RNA chaperone Hfq influences central carbon metabolism and the symbiotic interaction with alfalfa. BMC Microbiol 2010; 10:71; PMID:20205931; http://dx.doi.org/10.1186/1471-2180-10-71

61. Wilms I, Möller P, Stock AM, Gurski R, Lai EM, Narberhaus F. Hfq influences multiple transport systems and virulence in the plant pathogen Agrobacterium tumefaciens. J Bacteriol 2012; 194:5209-17; PMID:22821981; http://dx.doi.org/10.1128/JB.00510-12

62. Vogel J, Luisi BF. Hfq and its constellation of RNA. Nat Rev Microbiol 2011; 9:578-89; PMID:21760622; http://dx.doi.org/10.1038/nrmicro2615

63. Møller T, Franch T, Højrup P, Keene DR, Bächinger HP, Brennan RG, Valentin-Hansen P. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol Cell 2002; 9:23-30; PMID:11804583; http://dx.doi.org/10.1016/S1097-2765(01)00436-1

64. Garfinkel DJ, Nester EW. *Agrobacterium tumefaciens* mutants affected in crown gall tumorigenesis and octopine catabolism. J Bacteriol 1980; 144:732-43; PMID:6253441

65. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 2008; 9:474; PMID:19014431; http://dx.doi.org/10.1186/1471-2105-9-474

66. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res 2010; 38:W695-9; PMID:20439314; http://dx.doi.org/10.1093/nar/gkq313

67. Vogel J, Wagner EG. Target identification of small noncoding RNAs in bacteria. Curr Opin Microbiol 2007; 10:262-70; PMID:17574901; http://dx.doi.org/10.1016/j.mib.2007.06.001

68. Lease RA, Cusick ME, Belfort M. Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. Proc Natl Acad Sci U S A 1998; 95:12456-61; PMID:9770507; http://dx.doi.org/10.1073/pnas.95.21.12456

69. Massé E, Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc Natl Acad Sci U S A 2002; 99:4620-5; PMID:11917098; http://dx.doi.org/10.1073/pnas.032066599

70. Guillier M, Gottesman S. The 5′ end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. Nucleic Acids Res 2008; 36:6781-94; PMID:18953042; http://dx.doi.org/10.1093/nar/gkn742

71. Pfeiffer V, Papenfort K, Lucchini S, Hinton JC, Vogel J. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat Struct Mol Biol 2009; 16:840-6; PMID:19620966; http://dx.doi.org/10.1038/nsmb.1631

72. Durand S, Storz G. Reprogramming of anaerobic metabolism by the FnrS small RNA. Mol Microbiol 2010; 75:1215-31; PMID:20070527; http://dx.doi.org/10.1111/j.1365-2958.2010.07044.x

73. Beisel CL, Storz G. The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli.* Mol Cell 2011; 41:286-97; PMID:21292161; http://dx.doi.org/10.1016/j.molcel.2010.12.027

74. Papenfort K, Said N, Welsink T, Lucchini S, Hinton JC, Vogel J. Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. Mol Microbiol 2009; 74:139-58; PMID:19732340; http://dx.doi.org/10.1111/j.1365-2958.2009.06857.x

75. Rice JB, Vanderpool CK. The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes. Nucleic Acids Res 2011; 39:3806-19; PMID:21245045; http://dx.doi.org/10.1093/nar/gkq1219

76. Fröhlich KS, Papenfort K, Berger AA, Vogel J. A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. Nucleic Acids Res 2012; 40:3623-40; PMID:22180532; http://dx.doi.org/10.1093/nar/gkr1156

77. Gottesman S, Storz G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb Perspect Biol 2011; 3:3; PMID:20980440; http://dx.doi.org/10.1101/cshperspect.a003798

78. Papenfort K, Vogel J. Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. Res Microbiol 2009; 160:278-87; PMID:19366629; http://dx.doi.org/10.1016/j.resmic.2009.03.004

79. Massé E, Vanderpool CK, Gottesman S. Effect of RyhB small RNA on global iron use in *Escherichia coli.* J Bacteriol 2005; 187:6962-71; PMID:16199566; http://dx.doi.org/10.1128/JB.187.20.6962-6971.2005

80. Prévost K, Salvail H, Desnoyers G, Jacques JF, Phaneuf E, Massé E. The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. Mol Microbiol 2007; 64:1260-73; PMID:17542919; http://dx.doi.org/10.1111/j.1365-2958.2007.05733.x

81. Desnoyers G, Morissette A, Prévost K, Massé E. Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. EMBO J 2009; 28:1551-61; PMID:19407815; http://dx.doi.org/10.1038/emboj.2009.116

82. Salvail H, Lanthier-Bourbonnais P, Sobota JM, Caza M, Benjamin JA, Mendieta ME, Lépine F, Dozois CM, Imlay J, Massé E. A small RNA promotes siderophore production through transcriptional and metabolic remodeling. Proc Natl Acad Sci U S A 2010; 107:15223-8; PMID:20696910; http://dx.doi.org/10.1073/pnas.1007805107

83. Guillier M, Gottesman S, Storz G. Modulating the outer membrane with small RNAs. Genes Dev 2006; 20:2338-48; PMID:16951250; http://dx.doi.org/10.1101/gad.1457506

84. Schlüter JP, Reinkensmeier J, Daschkey S, Evguenieva-Hackenberg E, Janssen S, Jänicke S, Becker JD, Giegerich R, Becker A. A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti.* BMC Genomics 2010; 11:245; PMID:20398411; http://dx.doi.org/10.1186/1471-2164-11-245

85. Shelp BJ, Bown AW, Faure D. Extracellular gamma-aminobutyrate mediates communication between plants and other organisms. Plant Physiol 2006; 142:1350-2; PMID:17151138; http://dx.doi.org/10.1104/pp.106.088955

86. Moréra S, Gueguen-Chaignon V, Raffoux A, Faure D. Cloning, purification, crystallization and preliminary X-ray analysis of a bacterial GABA receptor with a Venus flytrap fold. Acta Crystallogr Sect F Struct Biol Cryst Commun 2008; 64:1153-5; PMID:19052373; http://dx.doi.org/10.1107/S1744309108036555

87. Yuan ZC, Haudecoeur E, Faure D, Kerr KF, Nester EW. Comparative transcriptome analysis of *Agrobacterium tumefaciens* in response to plant signal salicylic acid, indole-3-acetic acid and gamma-amino butyric acid reveals signalling cross-talk and *Agrobacterium*--plant co-evolution. Cell Microbiol 2008; 10:2339-54; PMID:18671824; http://dx.doi.org/10.1111/j.1462-5822.2008.01215.x

88. Planamente S, Vigouroux A, Mondy S, Nicaise M, Faure D, Moréra S. A conserved mechanism of GABA binding and antagonism is revealed by structure-function analysis of the periplasmic binding protein Atu2422 in *Agrobacterium tumefaciens.* J Biol Chem 2010; 285:30294-303; PMID:20630861; http://dx.doi.org/10.1074/jbc.M110.140715

89. Fröhlich KS, Vogel J. Activation of gene expression by small RNA. Curr Opin Microbiol 2009; 12:674-82; PMID:19880344; http://dx.doi.org/10.1016/j.mib.2009.09.009

90. Majdalani N, Hernandez D, Gottesman S. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. Mol Microbiol 2002; 46:813-26; PMID:12410838; http://dx.doi.org/10.1046/j.1365-2958.2002.03203.x

91. Sledjeski D, Gottesman S. A small RNA acts as an antisilencer of the H-NS-silenced *rcsA* gene of *Escherichia coli.* Proc Natl Acad Sci U S A 1995; 92:2003-7; PMID:7534408; http://dx.doi.org/10.1073/pnas.92.6.2003

92. Sledjeski DD, Gupta A, Gottesman S. The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli.* EMBO J 1996; 15:3993-4000; PMID:8670904

93. Fröhlich KS, Papenfort K, Fekete A, Vogel J. A small RNA activates CFA synthase by isoform-specific mRNA stabilization. EMBO J 2013; 32:2963-79; PMID:24141880; http://dx.doi.org/10.1038/emboj.2013.222

94. Mandin P, Gottesman S. Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. EMBO J 2010; 29:3094-107; PMID:20683441; http://dx.doi.org/10.1038/emboj.2010.179

95. Hilbers F, Eggers R, Pradela K, Friedrich K, Herkenhoff-Hesselmann B, Becker E, Deckers-Hebestreit G. Subunit δ is the key player for assembly of the H(+)-translocating unit of Escherichia coli F(O)F1 ATP synthase. J Biol Chem 2013; 288:25880-94; PMID:23864656; http://dx.doi.org/10.1074/jbc.M113.484675

96. Norrander J, Kempe T, Messing J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 1983; 26:101-6; PMID:6323249; http://dx.doi.org/10.1016/0378-1119(83)90040-9

97. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976; 72:248-54; PMID:942051; http://dx.doi.org/10.1016/0003-2697(76)90527-3

98. Wessel D, Flügge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem 1984; 138:141-3; PMID:6731838; http://dx.doi.org/10.1016/0003-2697(84)90782-6

99. Bandow JE, Baker JD, Berth M, Painter C, Sepulveda OJ, Clark KA, Kilty I, VanBogelen RA. Improved image analysis workflow for 2-D gels enables large-scale 2-D gel-based proteomics studies--COPD biomarker discovery study. Proteomics 2008; 8:3030-41; PMID:18618493; http://dx.doi.org/10.1002/pmic.200701184

100. Klüsener S, Hacker S, Tsai YL, Bandow JE, Gust R, Lai EM, Narberhaus F. Proteomic and transcriptomic characterization of a virulence-deficient phosphatidylcholine-negative *Agrobacterium tumefaciens* mutant. Mol Genet Genomics 2010; 283:575-89; PMID:20437057; http://dx.doi.org/10.1007/s00438-010-0542-7

101. Aiba H, Adhya S, de Crombrugghe B. Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. J Biol Chem 1981; 256:11905-10; PMID:6271763

102. Brantl S, Wagner EG. Antisense RNA-mediated transcriptional attenuation occurs faster than stable antisense/target RNA pairing: an *in vitro* study of plasmid pIP501. EMBO J 1994; 13:3599-607; PMID:7520390

103. Hartz D, McPheeters DS, Traut R, Gold L. Extension inhibition analysis of translation initiation complexes. Methods Enzymol 1988; 164:419-25; PMID:2468068; http://dx.doi.org/10.1016/S0076-6879(88)64058-4

104. Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, et al. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic Acids Res 2014; PMID:24838564; http://dx.doi:10.1093/nar/gku359

105. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003; 31:3406-15; PMID:12824337; http://dx.doi.org/10.1093/nar/gkg595

6 CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains

Patrick R. Wright†, Jens Georg†, Martin Mann†, Dragos A. Sorescu, Andreas S. Richter, Steffen Lott, Robert Kleinkauf, Wolfgang R. Hess and Rolf Backofen (2014) **Nucleic Acids Research**, 42, W119-W123.

† Shared first authors

DOI: 10.1093/nar/gku359

Personal contribution

I conceived and wrote the manuscript for this publication. Furthermore, I was centrally involved in the design of the webserver for both CopraRNA and IntaRNA. I also implemented a new backend for the IntaRNA webserver, which supports the enhanced output described in the paper.

Patrick R. Wright

The following co-authors confirm the above stated contribution.

Prof. Dr. Rolf Backofen

Dr. Martin Mann

Published online 16 May 2014 *Nucleic Acids Research, 2014, Vol. 42, Web Server issue* ***W119–W123***
doi: 10.1093/nar/gku359

CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains

Patrick R. Wright[1,2,†], Jens Georg[2,†], Martin Mann[1,†], Dragos A. Sorescu[1], Andreas S. Richter[1,3], Steffen Lott[2], Robert Kleinkauf[1], Wolfgang R. Hess[2] and Rolf Backofen[1,4,5,6,*]

[1]Bioinformatics Group, Department of Computer Science, Albert-Ludwigs-University Freiburg, Georges-Köhler-Allee 106, D-79110 Freiburg, Germany, [2]Genetics and Experimental Bioinformatics, Faculty of Biology, Schänzlestr. 1, D-79104 Freiburg, Germany, [3]Max Planck Institute of Immunobiology and Epigenetics, Stübeweg 51, D-79108 Freiburg, Germany, [4]BIOSS Centre for Biological Signalling Studies, Cluster of Excellence, Albert-Ludwigs-University Freiburg, Germany, [5]Center for non-coding RNA in Technology and Health, University of Copenhagen, Gronnegardsvej 3, DK-1870 Frederiksberg C, Denmark and [6]ZBSA Centre for Biological Systems Analysis, Albert-Ludwigs-University Freiburg, Habsburgerstr. 49, D-79104 Freiburg, Germany

Received January 20, 2014; Revised April 3, 2014; Accepted April 15, 2014

ABSTRACT

CopraRNA (Comparative prediction algorithm for small RNA targets) is the most recent asset to the Freiburg RNA Tools webserver. It incorporates and extends the functionality of the existing tool IntaRNA (Interacting RNAs) in order to predict targets, interaction domains and consequently the regulatory networks of bacterial small RNA molecules. The CopraRNA prediction results are accompanied by extensive postprocessing methods such as functional enrichment analysis and visualization of interacting regions. Here, we introduce the functionality of the CopraRNA and IntaRNA webservers and give detailed explanations on their postprocessing functionalities. Both tools are freely accessible at http://rna.informatik.uni-freiburg.de.

INTRODUCTION

In recent years, bacterial small RNAs (sRNAs) have proven to be potent, versatile and important regulators of prokaryotic gene expression (1,2). Furthermore, they are extremely abundant in various prokaryotic genomes (3–7) and due to novel experimental (6,8,9) and computational (10–12) methods on the genomic scale, biologists are struggling with ever increasing magnitudes of sRNA data that can, in many cases, only be harnessed by bioinformatics analyses (i.e. target predictions), preceding wetlab verifications. To make analysis methods accessible to a broad audience, graphical user interfaces (GUIs) are indispensable. Offering such interfaces in a web browser based manner has proven to be useful and intuitive to many users in the past (13–16). The Freiburg RNA Tools webserver aims at supplying an easy to use, free and comprehensive web resource for RNA analysis, also for non-adept users.

Several sRNA target prediction algorithms have been developed in the past (17), and many of them are available as webservers (14,18–21). Here, we highlight that CopraRNA (Comparative prediction algorithm for small RNA targets) (22) and IntaRNA (Interacting RNAs) (23) not only produce more than sound results but also supply postprocessing that greatly aids in the interpretation and evaluation of the results. The tools are accompanied by extensive help pages, and direct help requests are rapidly answered. The results can be viewed in the browser and downloaded for further local analysis or archiving. Furthermore, the source code for both tools is available for download on the Freiburg RNA software page at http://www.bioinf.uni-freiburg.de/Software/.

CopraRNA AND IntaRNA

While CopraRNA is a comparative method that constructs a combined sRNA target prediction for a set of given organisms, IntaRNA predicts interactions in single organisms. An exemplary workflow incorporating both tools is given in Figure 1. Employing a statistical model, CopraRNA computes whole genome target predictions by combining whole genome IntaRNA target screens for homologous sRNA sequences from distinct organisms. Individual evolutionary distances between the organisms and the statistical dependencies in the data are accounted for and are corrected within the workflow of the algorithm. IntaRNA predicts interacting regions between two RNA molecules by incorporating the accessibility of both interaction sites and the presence of a seed interaction; both features are commonly observed in sRNA–mRNA interactions (24). IntaRNA, un-

[*]To whom correspondence should be addressed. Tel: +49 761 2037460; Fax: +49 761 2037462; Email: backofen@informatik.uni-freiburg.de
[†]These authors contributed equally to this work.

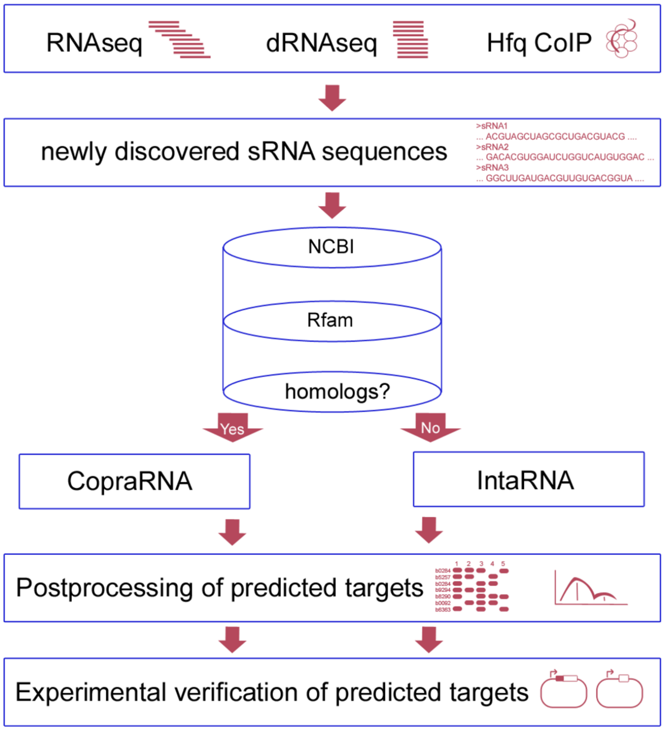

RNAseq
dRNAseq
Hfq CoIP
newly discovered sRNA sequences
>sRNA1
... ACGUAGCUAGCGCUGACGUACG ....
>sRNA2
... GACACGUGGAUCUGGUCAUGUGGAC ...
>sRNA3
... GGCUUGAUGACGUUGUGACGGUA ....
NCBI
Rfam
homologs?
Yes
No
CopraRNA
IntaRNA
Postprocessing of predicted targets
1
2
3
4
5
b0284
b5257
b0284
b9294
b8290
b0092
b6363
Experimental verification of predicted targets

Figure 1. sRNA identification and classification workflow incorporating CopraRNA or IntaRNA. The first box mentions selected experiments that have aided in sRNA identification, i.e. RNAseq (8), dRNAseq (6) or Hfq co-immunoprecipitation (CoIP) (9). The cylinder represents databases that can be queried while looking for sRNA homologs. Examples are NCBI (BLAST) (26) or Rfam (27). The next step is the execution of the actual sRNA target prediction depending on presence of sRNA homologs (CopraRNA) or absence of sRNA homologs (IntaRNA). The final two stages consist of postprocessing and selection of candidates for experimental verification, e.g. by a GFP reporter system (32).

like CopraRNA, can also be applied to non-whole genome screens using smaller sets of RNA molecules as input. Thus, it is also applicable to RNA–RNA interaction prediction for eukaryotic systems (25).

INPUT AND OUTPUT

Input data must be supplied in FASTA format. For CopraRNA, the FASTA file should represent three or more homologous sRNA sequences from distinct organisms. Homologous sRNA sequences may be retrieved from databases such as NCBI via BLAST (26) or from Rfam (27). While only three input sequences are mandatory, we suggest using at least five if available. CopraRNA requires for each sequence, a RefSeq ID of its affiliated organism as FASTA header (see Figure 3A, top left for an example). If several RefSeq IDs correspond to replicons of one organism, any one of these IDs may be supplied. A maximum of eight input organisms is possible. One of these species must be selected as central reference (organism of interest) for postprocessing and annotation.

Currently, ∼2700 organisms are available for CopraRNA and IntaRNA whole genome target predictions and the list is updated on a monthly basis. As previously mentioned, IntaRNA can also compute interactions for smaller sets of RNAs. In this case, the user may supply two FASTA files. For these, all pairwise interactions are computed. Suggested standard parameters for IntaRNA are a seed length (p) of 7, a target folding window size (w) of 150 and a maximum base pair distance (L) of 100 (28).

Both webservers provide the top 100 predictions of the respective methods as primary result table. Furthermore, the core results of the algorithms are accompanied by extensive postprocessing that aids interpretation and condensation of the result tables. For whole genome target predictions, CopraRNA and IntaRNA include automatic functional enrichment (29) of the top predicted targets and visualization of putative interacting regions within the sRNA and the mRNA. As a new feature of the webserver, the functionally enriched terms are represented within a heatmap, allowing 'at a glance' conclusions for target networks (see Figure 2 for an example). These results can guide the user while constructing functional networks and characterizing target binding mechanisms of a given sRNA. For users interested in the entire results, the corresponding job is available for download as compressed archive. Sample input and output pages for CopraRNA are displayed in Figure 3. Both tools' source code is also available for download and local installation from the Freiburg RNA webserver download page.

METHODS

CopraRNA utilizes IntaRNA to calculate single organism whole genome target predictions. IntaRNA predictions are computed for each sRNA-organism pair participating in the analysis. These individual predictions are the basis for the comparative model. In order to combine target predictions for homologous genes from distinct organisms, IntaRNA p-values are computed from the IntaRNA energy scores for each putative target with an energy score ≤ 0. Transforming energy scores to p-values is achieved by fitting generalized extreme value distributions to the IntaRNA energy scores. Using the resulting equations for each individual whole genome target prediction, p-values can be calculated for each putative target. In the following, the DomClust (30) algorithm is applied in order to cluster homologous genes. The clustering is based on the amino acid sequences of the organisms' protein coding genes. These clusters are then used to calculate a combined CopraRNA p-value for each cluster of homologous genes by employing Hartung's method for the combination of dependent p-values (31). Conveniently, it not only allows to account for the overall dependency within the data, but also incorporates the possibility to weight individual p-values. This is important, as the organisms participating in the analysis can usually not be regarded as equidistant. Closer organisms are consequently down weighted. Excessive influence of outliers is corrected for by applying a root function to the weights. The final set of CopraRNA p-values is employed for q-value calculation. The q-values give an estimate of the false discovery rate of the target prediction. More detailed algorithmic explanations on CopraRNA and IntaRNA are given in the original publications (22,23).

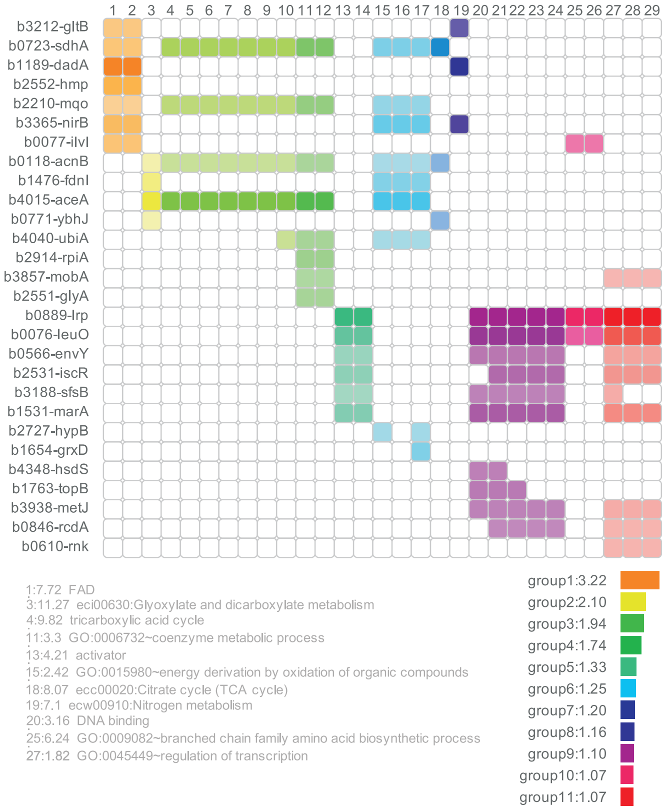


Figure 2. The CopraRNA heatmap shows the targets with a p-value ≤ 0.01 (for IntaRNA the top 50 predicted targets are subjected to the initial functional enrichment), which have homologs in the organism of interest and are functionally enriched. All members of clusters with a DAVID enrichment score ≥ 1.0 are shown in a specific color. Each row represents a gene and each column a specific functional term. If the gene can be assigned to a term, the corresponding square is colored. If no assignment was made, the square remains white. Closely related terms are assigned to a cluster and have the same color. The opacity of the color depends on the p-value of the CopraRNA prediction. A more intense color represents a more significant p-value. The 'fold enrichment' is given in front of the term descriptions. It represents the enrichment of a term in the prediction group in relation to the whole prediction background (e.g. a term with an enrichment of 10 contains 10 times more genes belonging to the respective term than the background). The enrichment scores give a measure of the biological significance of the cluster. The DAVID enrichment score for a cluster is the log transformed geometric mean of all enrichment p-values from the terms belonging to the respective cluster. A higher score represents a more statistically significant enrichment. The individual p-values for the terms are calculated by a modified Fisher's exact test. The length of the bars next to the groups of enriched genes corresponds to the size of the enrichment score. The publication on the DAVID webserver suggests to investigate clusters with an enrichment score of ≥ 1.3 while also pointing out that clusters with lower enrichment scores must not necessarily be discarded and may also contain useful information (33). This specific heatmap represents the enrichment output for the enterobacterial (here *Escherichia coli*) sRNA FnrS. Due to space reasons only one term for each cluster is shown.

POSTPROCESSING AND PREDICTION QUALITY ESTIMATION

The benchmarking of CopraRNA showed that some predictions are more reliable (e.g. GcvB, RyhB, FnrS) than others (e.g. ArcZ) (22). On behalf of a reduced experimental (32) workload it is preferable to have a measure for the reliability of each individual prediction. Here the q-value and the postprocessing outputs provide guidance. A strong functional enrichment signature, pointing to a specific group of genes or a specific pathway, has proven to be a reliable signal for a meaningful prediction. However, functional enrichments are not always present. This may be due to low prediction quality, but it can also be caused by a lack of annotation for the organism of interest or its absence in the DAVID knowledge base (33).

In these cases the user may opt to choose the organism with the best available annotation as organism of interest. If this proves ineffective, the user should resort to the q-value distribution and the interaction domain plots. A slowly growing q-value, i.e. a relatively high number of predictions with a q-value ≤ 0.5, is a hallmark of a more reliable prediction, especially if the interaction plots show distinct clustered interaction regions for the sRNA and mRNA homologs. A random distribution of the interaction sites in the mRNAs and/or sRNA homologs argues against a reliable prediction.

JOB ARCHIVING

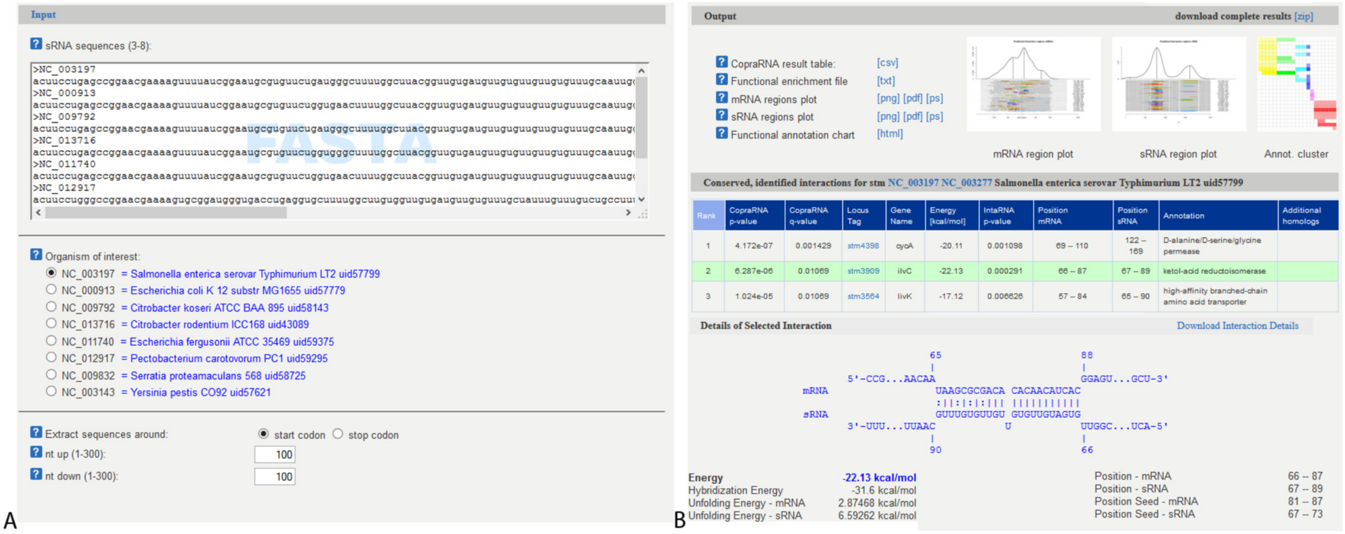
Upon submission, a unique ID, which is only known by the submitting user, is automatically assigned to each job. This ID can be used to recall the results of a specific job at any time within the storage period. The Freiburg RNA webserver stores all computed results for 30 days. Within this time, selected results or the entire job directory may be downloaded for local archiving by the user. Online archiving within the 30 day period is aided by the possibility of setting job specific descriptions.

PRIOR APPLICATION AND EVALUATION

The predictive performance of CopraRNA and IntaRNA was previously evaluated on an extensive benchmarking dataset of 101 experimentally verified sRNA and target pairs from 18 enterobacterial sRNAs (22). They were compared to each other and to RNApredator (19) and TargetRNA (18). Both tools from the Freiburg RNA webserver outperformed the other tools in predictive accuracy. Furthermore, CopraRNA was compared to experimental target prediction by micro arrays. Strikingly, it showed similar predictive quality with respect to the abundance of correctly predicted targets (22). From the CopraRNA benchmark predictions, 23 previously unreported, putative sRNA targets were selected for experimental verification. From these, 17 were verified (22). This represents a success rate of ∼74%. CopraRNA has also been successfully applied in studies on non-enterobacterial species. These include investigations of the sRNAs PsrR1 from *Synechocystis sp.* PCC6803 and AbcR1 from *Agrobacterium tumefaciens* (unpublished data). Beside many other studies, computational predictions with IntaRNA enabled the identification of two novel targets of the cyanobacterial sRNA Yfr1 (34) and aided in finding that the archaeal sRNA162 targets both *cis*- and *trans*-encoded mRNAs via two distinct domains (35).

IMPLEMENTATION

The Freiburg RNA webserver is based on Apache Tomcat Java Server Pages (JSP) to enable a high server-side perfor-



Input

? sRNA sequences (3-8):

```
>NC_003197
acuuccugagccggaacgaaaaguuuuaucggaaugcguguucugaugggcuuuuggcuuacgguugugauguuguguuguuguguuugcaauug
>NC_000913
acuuccugagccggaacgaaaaguuuuaucggaaugcguguucuggugaacuuuuggcuuacgguugugauguuguguuguuguguuugcaauug
>NC_009792
acuuccugagccggaacgaaaaguuuuaucggaaugcguguucugaugggcuuuuggcuuacgguugugauguuguguuguuguguuugcaauug
>NC_013716
acuuccugagccggaacgaaaaguuuuaucggaaugcguguucuggugggcuuuuggcuuacgguugugauguuguguuguuguguuugcaauug
>NC_011740
acuuccugagccggaacgaaaaguuuuaucggaaugcguguucuggugaacuuuuggcuuacgguugugauguuguguuguuguguuugcaauug
>NC_012917
acuuccugggccggaacgaaaagugcggaugggugaccugaggugcuuuuggcuugugguugugauguuguguuugcuauuuguuugucugccuu
```

? Organism of interest:

- ● NC_003197 = Salmonella enterica serovar Typhimurium LT2 uid57799
- ○ NC_000913 = Escherichia coli K 12 substr MG1655 uid57779
- ○ NC_009792 = Citrobacter koseri ATCC BAA 895 uid58143
- ○ NC_013716 = Citrobacter rodentium ICC168 uid43089
- ○ NC_011740 = Escherichia fergusonii ATCC 35469 uid59375
- ○ NC_012917 = Pectobacterium carotovorum PC1 uid59295
- ○ NC_009832 = Serratia proteamaculans 568 uid58725
- ○ NC_003143 = Yersinia pestis CO92 uid57621

? Extract sequences around: ● start codon ○ stop codon

? nt up (1-300): 100

? nt down (1-300): 100

Output download complete results [zip]

? CopraRNA result table: [csv]
? Functional enrichment file [txt]
? mRNA regions plot [png] [pdf] [ps]
? sRNA regions plot [png] [pdf] [ps]
? Functional annotation chart [html]

Conserved, identified interactions for stm NC_003197 NC_003277 Salmonella enterica serovar Typhimurium LT2 uid57799

Rank	CopraRNA p-value	CopraRNA q-value	Locus Tag	Gene Name	Energy [kcal/mol]	IntaRNA p-value	Position mRNA	Position sRNA	Annotation	Additional homologs
1	4.172e-07	0.001429	stm4398	cycA	-20.11	0.001098	69 -- 110	122 -- 169	D-alanine/D-serine/glycine permease	
2	6.287e-06	0.01069	stm3909	ilvC	-22.13	0.000291	66 -- 87	67 -- 89	ketol-acid reductoisomerase	
3	1.024e-05	0.01069	stm3564	livK	-17.12	0.006626	57 -- 84	65 -- 90	high-affinity branched-chain amino acid transporter	

Details of Selected Interaction Download Interaction Details

```
                   65                      88
                   |                       |
        5'-CCG...AACAA                     GGAGU...GCU-3'
mRNA               UAAGCGCGACA  CACAACAUCAC
                   :||:|:|:|||  |||||||||||
sRNA               GUUUGUGUUGU  GUGUUGUAGUG
        3'-UUU...UUAAC            U        UUGGC...UCA-5'
                   |                       |
                   90                      66
```

Energy	-22.13 kcal/mol	Position - mRNA	66 -- 87
Hybridization Energy	-31.6 kcal/mol	Position - sRNA	67 -- 89
Unfolding Energy - mRNA	2.87468 kcal/mol	Position Seed - mRNA	81 -- 87
Unfolding Energy - sRNA	6.59262 kcal/mol	Position Seed - sRNA	67 -- 73

Figure 3. CopraRNA webserver input (**A**) and output (**B**) page for the sRNA GcvB. The FASTA file may be pasted or uploaded to the webserver. Upon insertion of the sequences, the webserver automatically displays the RefSeq IDs' organism affiliations (blue text in (A)). The output page contains a visualization of the primary result table, the interaction as predicted by IntaRNA and the interacting region plots within the sRNA and mRNA. Furthermore, the functional enrichment is visualized as interactive heatmap.

mance for input validation, job execution and retrieval, and dedicated pre- and postprocessing. Javascripting is used to provide an intuitive and interactive user interface on the client side. The tools provided by the Freiburg RNA webserver are run on a dedicated computing cluster with up to 480 CPUs, depending on the workload. Jobs are automatically queued and started via Sun Grid Engine to ensure a balanced and fast job processing given the varying execution requirements of the different tools provided. An automatic emailing system informs the user upon job completion if an email address (optional) is provided upon submission.

ACKNOWLEDGMENTS

We thank Sita J. Saunders for helping with the paper formatting.

Authors' contributions. Conceived the methods and analyses: P.R.W., J.G., W.R.H., M.M., R.B., D.A.S., A.S.R. and S.L. Technical assistance: D.A.S., M.M. and R.K. Designed the webserver: M.M., D.A.S. and P.R.W. Planned and wrote the paper: P.R.W., J.G., M.M. and R.B. All authors read and revised the final manuscript.

FUNDING

Deutsche Forschungsgemeinschaft Focus Program Sensory and Regulatory RNAs in Prokaryotes [SPP1258]; Bundesministerium fuer Bildung und Forschung (BMBF) [0316165 to W.R.H. and R.B.]; Excellence Initiative of the German Federal and State Governments [EXC 294 to R.B.].

Conflict of interest statement. None declared.

REFERENCES

1. Storz,G., Vogel,J. and Wassarman,K.M. (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell*, **43**, 880–891.
2. Gottesman,S. and Storz,G. (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspect. Biol.*, **3**, a003798.
3. Dugar,G., Herbig,A., Forstner,K.U., Heidrich,N., Reinhardt,R., Nieselt,K. and Sharma,C.M. (2013) High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS Genet.*, **9**, e1003495.
4. Mitschke,J., Georg,J., Scholz,I., Sharma,C.M., Dienst,D., Bantscheff,J., Voß,B., Steglich,C., Wilde,A., Vogel,J. *et al.* (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 2124–2129.
5. Georg,J. and Hess,W.R. (2011) Regulatory RNAs in cyanobacteria: developmental decisions, stress responses and a plethora of chromosomally encoded cis-antisense RNAs. *Biol. Chem.*, **392**, 291–297.
6. Sharma,C.M., Hoffmann,S., Darfeuille,F., Reignier,J., Findeiß,S., Sittka,A., Chabas,S., Reiche,K., Hackermüller,J., Reinhardt,R. *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, **464**, 250–255.
7. Kroger,C., Colgan,A., Srikumar,S., Handler,K., Sivasankaran,S.K., Hammarlof,D.L., Canals,R., Grissom,J.E., Conway,T., Hokamp,K. *et al.* (2013) An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe*, **14**, 683–695.
8. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
9. Zhang,A., Wassarman,K.M., Rosenow,C., Tjaden,B.C., Storz,G. and Gottesman,S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.*, **50**, 1111–1124.
10. Voß,B., Georg,J., Schön,V., Ude,S. and Hess,W.R. (2009) Biocomputational prediction of non-coding RNAs in model cyanobacteria. *BMC Genomics*, **10**, 123.
11. Babski,J., Tjaden,B., Voss,B., Jellen-Ritter,A., Marchfelder,A., Hess,W.R. and Soppa,J. (2011) Bioinformatic prediction and experimental verification of sRNAs in the haloarchaeon Haloferax volcanii. *RNA Biol.*, **8**, 806–816.
12. Amman,F., Wolfinger,M.T., Lorenz,R., Hofacker,I.L., Stadler,P.F. and Findeiss,S. (2014) TSSAR: TSS annotation regime for dRNA-seq data. *BMC Bioinformatics*, **15**, 89.
13. Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
14. Smith,C., Heyne,S., Richter,A.S., Will,S. and Backofen,R. (2010) Freiburg RNA Tools: a web server integrating IntaRNA, ExpaRNA

and LocARNA. *Nucleic Acids Res.*, **38**(Web Server issue), W373–W377.

15. Sorescu,D.A., Möhl,M., Mann,M., Backofen,R. and Will,S. (2012) CARNA—alignment of RNA structure ensembles. *Nucleic Acids Res*, **40**, W49–W53.
16. Lange,S.J., Alkhnbashi,O.S., Rose,D., Will,S. and Backofen,R. (2013) CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res.*, **41**, 8034-8044 .
17. Backofen,R. and Hess,W.R. (2010) Computational prediction of sRNAs and their targets in bacteria. *RNA Biol.*, **7**, 33–42.
18. Tjaden,B., Goodwin,S.S., Opdyke,J.A., Guillier,M., Fu,D.X., Gottesman,S. and Storz,G. (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res.*, **34**, 2791–2802.
19. Eggenhofer,F., Tafer,H., Stadler,P.F. and Hofacker,I.L. (2011) RNApredator: fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res.*, **39**(Web Server issue), W149–W154.
20. Ying,X., Cao,Y., Wu,J., Liu,Q., Cha,L. and Li,W. (2011) sTarPicker: a method for efficient prediction of bacterial sRNA targets based on a two-step model for hybridization. *PLoS One*, **6**, e22705.
21. Cao,Y., Zhao,Y., Cha,L., Ying,X., Wang,L., Shao,N. and Li,W. (2009) sRNATarget: a web server for prediction of bacterial sRNA targets. *Bioinformation*, **3**, 364–366.
22. Wright,P.R., Richter,A.S., Papenfort,K., Mann,M., Vogel,J., Hess,W.R., Backofen,R. and Georg,J. (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E3487-E3496.
23. Busch,A., Richter,A.S. and Backofen,R. (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, **24**, 2849–2856.
24. Richter,A.S. and Backofen,R. (2012) Accessibility and conservation: general features of bacterial small RNA-mRNA interactions? *RNA Biol.*, **9**, 954–965.
25. Starczynowski,D.T., Morin,R., McPherson,A., Lam,J., Chari,R., Wegrzyn,J., Kuchenbauer,F., Hirst,M., Tohyama,K., Humphries,R.K. *et al.* (2011) Genome-wide identification of human microRNAs located in leukemia-associated genomic alterations. *Blood*, **117**, 595–607.
26. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
27. Burge,S.W., Daub,J., Eberhardt,R., Tate,J., Barquist,L., Nawrocki,E.P., Eddy,S.R., Gardner,P.P. and Bateman,A. (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.*, **41**(Database issue), D226–D232.
28. Lange,S.J., Maticzka,D., Möhl,M., Gagnon,J.N., Brown,C.M. and Backofen,R. (2012) Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic Acids Res.*, **40**, 5215–5226.
29. Jiao,X., Sherman,B.T., Huang,D.W., Stephens,R., Baseler,M.W., Lane,H.C. and Lempicki,R.A. (2012) DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, **28**, 1805–1806.
30. Uchiyama,I. (2006) Hierarchical clustering algorithm for comprehensive orthologous-domain classification in multiple genomes. *Nucleic Acids Res.*, **34**, 647–658.
31. Hartung,J. (1999) A note on combining dependent tests of significance. *Biom. J.*, **41**, 849–855.
32. Corcoran,C.P., Podkaminski,D., Papenfort,K., Urban,J.H., Hinton,J.C.D. and Vogel,J. (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol. Microbiol.*, **84**, 428–445.
33. Huang,D.W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
34. Richter,A.S., Schleberger,C., Backofen,R. and Steglich,C. (2010) Seed-based IntaRNA prediction combined with GFP-reporter system identifies mRNA targets of the small RNA Yfr1. *Bioinformatics*, **26**, 1–5.
35. Jäger,D., Pernitzsch,S.R., Richter,A.S., Backofen,R., Sharma,C.M. and Schmitz,R.A. (2012) An archaeal sRNA targeting *cis*- and *trans*-encoded mRNAs via two distinct domains. *Nucleic Acids Res.*, **40**, 10964–10979.

7 A stress-induced small RNA modulates alpha-rhizobial cell cycle progression

Marta Robledo, Benjamin Frage, **Patrick R. Wright** and Anke Becker (2015) **PLOS Genetics**, 11, e1005153.
DOI: 10.1371/journal.pgen.1005153

Personal contribution

In this project, I was involved in performing the CopraRNA in silico analyses for the small RNA EcpR1. Furthermore, I performed functional enrichment analyses on the predictions returned for EcpR1.

Patrick R. Wright

The following co-authors confirm the above stated contribution.

Prof. Dr. Anke Becker

Dr. Marta Robledo

RESEARCH ARTICLE

A Stress-Induced Small RNA Modulates Alpha-Rhizobial Cell Cycle Progression

Marta Robledo[1], Benjamin Frage[1], Patrick R. Wright[2], Anke Becker[1]*

1 LOEWE Center for Synthetic Microbiology and Faculty of Biology, Philipps-University Marburg, Marburg, Germany, **2** Bioinformatics Group, Department of Computer Science, Albert-Ludwigs-University Freiburg, Freiburg, Germany

* anke.becker@synmikro.uni-marburg.de

Citation: Robledo M, Frage B, Wright PR, Becker A (2015) A Stress-Induced Small RNA Modulates Alpha-Rhizobial Cell Cycle Progression. PLoS Genet 11(4): e1005153. doi:10.1371/journal.pgen.1005153

Editor: Patrick H. Viollier, University of Geneva Medical School, SWITZERLAND

Received: August 16, 2014

Accepted: March 18, 2015

Published: April 29, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files except for the transcriptome data, which are available at ArrayExpress (Accession no. E-MTAB-3389).

Funding: We acknowledge financial support from the Postdoctoral Program of the Alexander von Humboldt Foundation (www.humboldt-foundation.de) to MR, the LOEWE program of the State of Hesse (https://wissenschaft.hessen.de/loewe) in the framework of the Center for Synthetic Microbiology (SYNMIKRO) to AB, and from the German Research Foundation (www.dfg.de) (SPP 1258: Be 2121/5-2) to AB. The funders had no role in study design, data collection

Abstract

Mechanisms adjusting replication initiation and cell cycle progression in response to environmental conditions are crucial for microbial survival. Functional characterization of the *trans*-encoded small non-coding RNA (*trans*-sRNA) EcpR1 in the plant-symbiotic alpha-proteobacterium *Sinorhizobium meliloti* revealed a role of this class of riboregulators in modulation of cell cycle regulation. EcpR1 is broadly conserved in at least five families of the Rhizobiales and is predicted to form a stable structure with two defined stem-loop domains. In *S. meliloti*, this *trans*-sRNA is encoded downstream of the *divK-pleD* operon. *ecpR1* belongs to the stringent response regulon, and its expression was induced by various stress factors and in stationary phase. Induced EcpR1 overproduction led to cell elongation and increased DNA content, while deletion of *ecpR1* resulted in reduced competitiveness. Computationally predicted EcpR1 targets were enriched with cell cycle-related mRNAs. Post-transcriptional repression of the cell cycle key regulatory genes *gcrA* and *dnaA* mediated by mRNA base-pairing with the strongly conserved loop 1 of EcpR1 was experimentally confirmed by two-plasmid differential gene expression assays and compensatory changes in sRNA and mRNA. Evidence is presented for EcpR1 promoting RNase E-dependent degradation of the *dnaA* mRNA. We propose that EcpR1 contributes to modulation of cell cycle regulation under detrimental conditions.

Author Summary

Microorganisms frequently encounter adverse conditions unfavorable for cell proliferation. They have evolved diverse mechanisms, including transcriptional control and targeted protein degradation, to adjust cell cycle progression in response to environmental cues. Non-coding RNAs are widespread regulators of various cellular processes in all domains of life. In prokaryotes, *trans*-encoded small non-coding RNAs (*trans*-sRNAs) contribute to a rapid cellular response to changing environments, but so far have not been directly related to cell cycle regulation. Here, we report the first example of a *trans*-sRNA (EcpR1) with two experimentally confirmed targets in the core of cell cycle regulation and demonstrate that in the plant-symbiotic alpha-proteobacterium *Sinorhizobium meliloti*

and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

the regulatory mechanism involves base-pairing of this sRNA with the *dnaA* and *gcrA* mRNAs. Most trans-sRNAs are restricted to closely related species, but the stress-induced EcpR1 is broadly conserved in the order of Rhizobiales suggesting an evolutionary advantage conferred by *ecpR1*. It broadens the functional diversity of prokaryotic sRNAs and adds a new regulatory level to the mechanisms that contribute to interlinking stress responses with the cell cycle machinery.

Introduction

Non-coding RNAs (ncRNAs) have shot to prominence as significant and ubiquitous regulators that are involved in the control of various cellular processes in most eukaryotic and prokaryotic organisms. Although the development of deep-sequencing technologies has allowed for the identification of an ever-growing number of ncRNAs the biological functions and regulatory mechanisms of the vast majority remain veiled. In eukaryotes, short-interfering RNAs (siRNA) and microRNAs (miRNAs) have emerged as a priority research area in biomedicine [1] since they control crucial cellular processes, such as cell development, differentiation and oncogenic transformation [2]. For instance, the miR-34 family mimics p53 activity, inducing cell-cycle arrest and apoptosis [3]. Plant ncRNAs have been reported to regulate stress adaptation and defence responses, but also cell differentiation, such as miR169 that was associated with nodule development in legumes [4,5]. In the fission yeast *Schizosaccharomyces pombe*, meiRNA plays a role in recognition of homologous chromosomes for pairing and thus is essential for progression of meiosis [6,7].

Prokaryotic *trans*-encoded small RNAs (*trans*-sRNAs) may be considered functional analogs of eukaryotic siRNAs and miRNAs in their ability to post-transcriptionally control gene expression by modulating mRNA translation and stability. The canonical regulatory mechanism of bacterial *trans*-sRNAs involves pairing with a single short binding site within the 5'-untranslated region (UTR) of the target mRNA, which results in formation of an sRNA-mRNA duplex blocking the ribosome binding site (RBS) and/or promoting degradation by RNases [8]. Expression of bacterial sRNAs is commonly stimulated under stress conditions and contributes to the rapid cellular response and adaptation to changing environments. The majority of functionally characterized bacterial sRNAs controls crucial physiological processes like metabolism, transport, chemotaxis, virulence, and quorum sensing [9].

Regulation of DNA replication and cell cycle progression in response to environmental cues is critical to ensure cell survival. Mechanisms involving small molecule-based signaling, protein-protein interactions or regulated proteolysis have been implicated with a delay of replication initiation or septum formation upon facing hostile factors [10]. It is tempting to speculate that *trans*-sRNA-mediated post-transcriptional regulation may also contribute to rapid adaptive stress responses of the cell cycle control circuit in bacteria.

The α-proteobacterium *Caulobacter crescentus* is an important model organism for studying cell cycle regulation. In this bacterium, replication is initiated only once per cell cycle [11,12]. This tight control and exact timing is governed by oscillating concentrations of at least three master regulators, DnaA, GcrA, and CtrA that coordinate the spatio-temporal pattern of phase-specific events ultimately leading to asymmetric cell division [13,14]. DnaA mediates replication initiation and activates *gcrA* expression. GcrA controls components of the replication and segregation machinery and finally induces expression of *ctrA*. CtrA blocks replication initiation by binding to the origin of replication and regulates more than 100 genes. Among these are genes involved in cell division, cell wall metabolism, and motility [15,16]. CtrA activation is driven by the essential CckA-ChpT phosphorelay, which further inactivates CpdR-mediated CtrA proteolysis by phosphorylating this response regulator. When activated by its principal kinase DivJ, DivK silences the CckA-ChpT relay through DivL, allowing for CtrA degradation and replication initiation. Subsequently, DivK is inactivated by dephosphorylation through its primary phosphatase PleC [17].

In the class of α-proteobacteria, several surveys of the non-coding RNome delivered a plethora of *trans*-sRNAs [18–20]. The most comprehensive inventories were performed for members of the *Rhizobiaceae* including *Sinorhizobium meliloti* [21,22]. *S. meliloti* exists either in a free-living lifestyle in the soil or in root nodule symbiosis with a leguminous host plant [23,24]. It has emerged as model organism to study adaptation to stress conditions and switching between complex lifestyles. The cell cycle of *C. crescentus* and free-living *S. meliloti* shows striking similarities that include initiation of replication only once per cell cycle and asymmetric cell division. In spite of species-specific rearrangements of the α-proteobacterial cell cycle regulon, a transcriptional analysis of synchronized *S. meliloti* cells has recently identified a conserved core of cell cycle regulated transcripts shared with *C. crescentus* [25] and confirmed previous computational comparisons of cell cycle-related genes in α-proteobacteria [26].

Taking advantage of the comprehensive data resource of *trans*-sRNAs in *S. meliloti* and related α-proteobacteria, we aimed at identifying riboregulators that post-transcriptionally affect bacterial cell cycle progression. Here, we report on the functional analysis of the stress-induced *trans*-sRNA EcpR1 that is conserved in several members of the Rhizobiales. We present evidence for EcpR1 negatively regulating *dnaA* and *gcrA* at the post-transcriptional level mediated by base-pairing between a strongly conserved loop of this sRNA and the target mRNAs. Our data suggests that EcpR1 contributes to a regulatory network connecting stress adaptation and cell cycle progression.

Results

EcpR1 target prediction shows enrichment of cell cycle-related genes

Hypothesizing that riboregulators affecting cell cycle control are more likely to be found among phylogenetically conserved *trans*-sRNAs we performed mRNA target predictions for 27 previously defined RNA families with members in at least two species [27] applying CopraRNA [28]. The predicted targets were screened for an enrichment of cell cycle-related genes. The CopraRNA algorithm considers base pairing strength, hybridization free energy and accessibility of the interaction sites, and integrates phylogenetic information to predict conserved sRNA-mRNA interactions. Many sRNAs base pair at the RBS, however, translation can also be blocked when the pairing region is located 50 or more nucleotides (nt) upstream the RBS or in the open reading frame [29,30]. As suggested by Wright et al. [28], predictions were therefore based on sequences 200 nt upstream and 100 nt downstream of the annotated start codons.

Targets predicted for the sRNA family established by the *S. meliloti trans*-sRNA SmelC291 show a significant enrichment (P-value = $2.5^{*}10^{-5}$) of cell cycle-related mRNAs (n = 7) among the top-ranked candidates ($P \leq 0.01$, n = 89; S1 Table) [27]. The 23 family members are broadly distributed among the Rhizobiales including members in the *Rhizobiaceae*, *Phyllobacteriaceae*, *Xanthobacteriaceae*, *Beijerinckaceae*, and *Hyphomicrobiaceae*. SmelC291, previously named SmrC10 or Sra33, was first identified by comparative genomic predictions of sRNAs [31] and confirmed by RNAseq [21]. In this study we renamed it EcpR1 (elongated cell phenotype RNA1) according to the phenotype induced by its overproduction (see below). In *S. meliloti*, *ecpR1* is located in the intergenic region between the *divK-pleD* operon coding for an essential cell cycle response regulator and a diguanylate cyclase, respectively [32] and *rpmG* encoding the 50S ribosomal protein L33 (Fig 1A). In the *Rhizobiacea*, this genomic locus is highly microsyntenic [27]. Northern blot hybridizations confirmed *ecpR1* expression from an independent transcription unit [33] and RNAseq coverage data suggested variants of different length with a dominant 101 nt sRNA [21] which is predicted to form a stable structure with two defined stem-loop domains, SL1 and SL2 (Fig 1A, S1A Fig). SL1 is strongly conserved and positions C_{16} to G_{36} (according to the numbering of EcpR1 nucleotides in Fig 1A) including the loop sequence are identical in all species with EcpR1 homologs analyzed by Reinkensmeier et al. [27]. The 3'-region harbors a putative Rho-independent terminator and 4 terminal U residues (S1A Fig).

In the *Rhizobiaceae*, *gcrA*, *dnaA*, and *pleC* mRNAs appeared among the five top predicted targets (positions 1, 3 and 5, respectively). Furthermore, the two *ftsZ* homologs *(ftsZ1* and *ftsZ2)*, *ctrA* and *minD* encoding a close homolog of the *Escherichia coli* cell division inhibitor [34] were in the top 40 list ($P<0.005$) of EcpR1 targets (S1 Table). Although there was less agreement with targets predicted in more distantly related members of the Rhizobiales, *gcrA* and *minD* mRNAs were also assessed as highly probable targets when predictions included *Mesorhizobium* strains belonging to the *Phyllobacteriaceae* ($P<0.001$) or members of the *Xanthobacteriaceae* ($P<0.007$). Finally, the *pleC* mRNA was still among the top target candidates ($P<0.0001$) when members of the *Xanthobacteriaceae*, *Beijerinckiaceaceae*, and *Hyphomicrobiaceae* were analyzed.

The GC-rich conserved region within SL1 of EcpR1 was predicted to base pair with all cell cycle-related target mRNA candidates (Fig 1A, S1B Fig). The interacting sequences predicted by CopraRNA were found in different positions of the *S. meliloti* mRNAs: for *gcrA*, a 13 nt stretch from position -109 to -95 relative to the start codon (S9D Fig); for *ctrA*, a 8 nt sequence from position -21 to -12 located close to the RBS; and for *pleC* and *minD*, discontinuous base-pairing over a 13 nt stretch overlapping the start codon and the *minC-minD* intergenic region, respectively (S7A–S7C Fig). The putative binding sites in the *dnaA* (S1C Fig; BS5) and *ftsZ* mRNAs (S7D Fig) map to positions about 60 to 70 nt downstream of the start codon. Additionally, the mRNA sequences ranging from the mapped *S. meliloti* transcriptional start site (TSS) [22] to 100 nt downstream of the annotated start codon were scanned for further sequences that may interact with EcpR1 applying IntaRNA [35]. This approach suggested three additional putative EcpR1 binding sites in the *dnaA* mRNA with $E < -10$ kcal/mol (S1C Fig): two at positions -140 and -70 relative to the AUG (BS1 and BS2), and a sequence overlapping the start codon region (BS3). The RNAup webserver [36] also identified these putative EcpR1 binding sites together with a sequence overlapping the RBS (BS4) (S1C Fig).

ecpR1 is expressed upon entry into stationary phase and under stress conditions

Microarray-based transcriptome profiling detected EcpR1 upon heat, cold, acidic, alkaline, salt, and oxidative stresses [21,37]. In the *S. meliloti* Rm2011 wild type, Northern blots revealed a dominant ~100 nt EcpR1 transcript and two less abundant larger variants corresponding to the prevalent 101 nt species, a 142 nt transcript, and the full length 171 nt variant deduced from the RNAseq data (Fig 1A and 1C; S1A Fig). In TY rich medium EcpR1 was barely detected in exponentially growing bacteria (OD_{600} of 0.2 to 0.9), and levels increased during early and late stationary phases (OD_{600} of 1.2 to 2.8) (Fig 1C, S2A Fig). The amount of EcpR1 also increased after shifting exponential phase cultures for one hour to 40°C, 20°C, or microoxic conditions, and after adding salt or hydrogen peroxide (~1.5 to 2 fold induction) (Fig 1C). qRT-PCR quantification of EcpR1 transcripts including the sequence region of the 101 nt variant even suggested higher induction levels (up to ~5-fold upon temperature upshift) (Fig 1C). EcpR1 levels

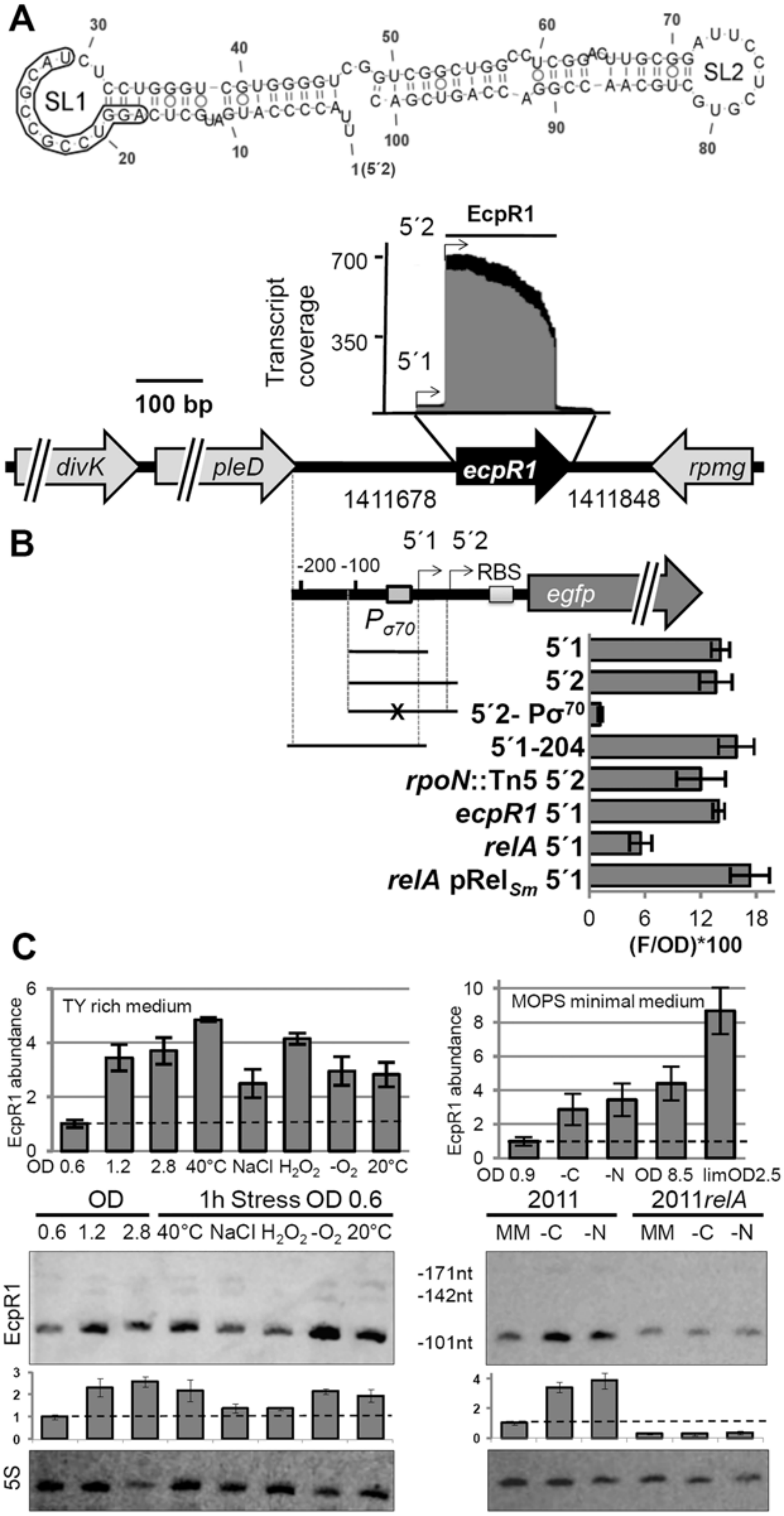


Fig 1. *ecpR1* genomic locus and transcriptional regulation. (A) Secondary structure of the dominant EcpR1 101 nt variant with a minimum free energy of -50.20 kcal/mol. Nucleotide positions relative to the second 5'-end are denoted. SL, stem loop domain. The 13 nt region predicted to bind the *gcrA* mRNA is boxed. Below, chromosomal region including the *ecpR1* gene and RNAseq coverage profile of the EcpR1 sRNA in *S. meliloti* Rm1021. Genome coordinates of the full length *ecpR1* variant are denoted. Black and

grey areas represent coverages from samples enriched for processed and primary transcripts, respectively [21]. Detected EcpR1 5'-ends are depicted by arrows and the dominant 101 nt EcpR1 variant used for structure prediction is marked by the bar. **(B)** Schematic representation of the fragments included in the *ecpR1* transcriptional fusions and fluorescence values of stationary phase Rm2011 wild type and derivative cells harbouring the indicated constructs: 5'1, pP*ecpR1*_5'1; 5'2, pP*ecpR1*_5'2; 5'2-Pσ70, pP*ecpR1*_5'2-Pσ70; 5'1–204, pP*ecpR1*_5'1–204. Specific activities were normalized to OD_{600} to yield fluorescence units per unit of optical density (F/OD). Shown are means and standard deviation values of at least three independent measurements of three transconjugants grown in six independent cultures. **(C)** qRT-PCR analysis and Northern blot detection of EcpR1 transcript abundance in Rm2011 and the *relA* mutant under different growth and stress conditions in TY (left) and MOPS minimal and MOPSlim medium (MM, right). 40°C, heat stress; NaCl, 0.4 mM sodium chloride (osmotic stress); H_2O_2, 10mM hydrogen peroxide (oxidative stress); $-O_2$, microoxic conditions; 20°C, cold stress; -C and -N, growth in MM until OD_{600} of 0.9 and then MM depleted for 1 hour for carbon or nitrogen. qRT-PCR values were normalized to the SMc01852 transcript and the levels of EcpR1 in Rm2011 growing in TY rich medium at OD_{600} of 0.6 (left) or MOPS minimal medium at OD_{600} of 0.9 (right, dashed line). Plots underneath the Northern blots represent relative hybridization signal intensities. The basal level of EcpR1 in Rm2011 growing in TY rich medium at OD_{600} of 0.6 or MOPS minimal medium at OD_{600} of 0.9 (right) has been normalized to 1 (dashed line) and the sRNA levels in other conditions have been correlated to this value. Mean results from three experiments are shown. Error bars indicate the standard deviation. Exposure times were optimized for each panel.

doi:10.1371/journal.pgen.1005153.g001

also increased when exponential phase cells growing in MOPS minimal medium were shifted to carbon or nitrogen depleted medium for one hour (~3.5-fold induction) (Fig 1C). Higher induction rates were observed in MOPS and nutrient-limited MOPS (MOPSlim) stationary phase cultures (up to ~8.5-fold, Fig 1C). Under these conditions, the stationary phase was reached at OD_{600} of 8.5 and 2.5, respectively. EcpR1 was not detected in total RNA isolated from 28 days old mature symbiotic nodules of *Medicago sativa* (S2B Fig).

RNAseq identified two distinct 5'-ends of the *ecpR1* mRNA varying by 29 nt [22] (Fig 1A). Although these 5'-ends were associated to σ^{70}- (ATTGAT-N17-CAATGC) (Fig 1B) and σ^{54}-type (AGGAAGG-AAAC-TTCCA) promoter motifs (S2C Fig), the alternative 5'2-end may either be generated by the activity of the putative σ^{54}-dependent promoter or by post-transcriptional processing of the EcpR1 primary transcript. To determine promoter activities associated with *ecpR1*, different DNA fragments from the *ecpR1* upstream region including up to 12 nt downstream of the TSS were fused to *egfp* in a replicative low copy plasmid (Fig 1B). Matching the results from the Northern hybridizations and qRT-PCR, the pP*ecpR1*_5'2 construct showed very low activities, just surpassing background fluorescence in the exponential growth phase of Rm2011 cultures in TY rich medium, while in stationary phase activities strongly increased (S2D Fig). Microscopy of Rm2011 single cells carrying pP*ecpR1*_5'2 showed that overall fluorescence homogeneously increased in stationary phase (S2E Fig), further confirming the growth phase-dependent pattern of *ecpR1* expression. All constructs including the σ^{70} promoter motif (pP*ecpR1*_5'1, pP*ecpR1*_5'2, and pP*ecpR1*_5'1–204) showed similar activities in stationary phase (Fig 1B). Mutations in the -10 region of the σ^{70}-type promoter abolished fluorescence activity of the reporter plasmid pP*ecpR1*_5'2-Pσ^{70} (Fig 1B) and EcpR1 was not detected by Northern hybridizations in stationary growing and oxygen depleted 2011Pσ^{70}*ecpR1* bacteria carrying these promoter mutations in the genome (S2F Fig). Furthermore, in stationary cultures an *rpoN* mutation did not reduce the reporter gene activity mediated by the pP*ecpR1*_5'2 construct including both putative promoters (Fig 1B). This suggests that the predicted σ^{54}-type promoter is nonfunctional under the conditions tested and implies that the prominent 5'-end of EcpR1 was probably generated by ribonucleolytic activity. *ecpR1* was not required for stimulation of *ecpR1* promoter activity in the stationary phase excluding a positive feedback involving the EcpR1 sRNA (Fig 1B). In trans overproduction of PleD or DivK, encoded upstream of *ecpR1* (Fig 1A), did not affect activity of any of the reporter gene constructs (S2G Fig).

Since the predicted promoter motifs provide no hints to extracytoplasmic function sigma factors being involved in stress-induced stimulation of *ecpR1* expression, we assayed the role of the stringent response alarmone ppGpp in regulation of *ecpR1*. Previously reported transcriptome data of cultures shifted to nitrogen or carbon starvation indicated a 20-fold and 4-fold increase in EcpR1 levels in the wild type and *relA* mutant, respectively [38]. Compared to the wild type, stimulation of *ecpR1* expression was reduced more than two-fold in a *relA* mutant that is unable to synthesize ppGpp and was fully restored by ectopic *relA* expression driven by the basal activity of the non-induced *lac* promoter [38] (Fig 1B). This result is in agreement with comparable levels of EcpR1 in the *relA* mutant under nutrient-sufficient and nitrogen- or carbon-limiting conditions as inferred from Northern hybridizations (Fig 1C) suggesting that EcpR1 is part of the stringent response regulon in *S. meliloti*.

Overexpression of *ecpR1* leads to cell cycle defects in several related α-proteobacteria

To study the biological function of EcpR1, growth and morphology phenotypes were monitored in *S. meliloti* either overexpressing *ecpR1* or lacking a functional copy of this sRNA gene.

IPTG-induced overexpression of *ecpR1* was mediated by construct pSKEcpR1$^+$ in strain Rm4011 carrying mutations that prevent background activity of the applied inducible expression system (see materials and methods). Northern hybridizations verified IPTG-driven overexpression of *ecpR1* from plasmid pSKEcpR1$^+$. Due to the overall stronger signals, the three less abundant EcpR1 variants matching the RNAseq data [21] were clearly detected in addition to the dominant 101 nt EcpR1 transcript (Figs 1A and 2A; S1A Fig, S2A Fig). IPTG-driven overexpression of the SmelC812 RNA gene from plasmid pSKControl$^+$ served as control in all *ecpR1* overexpression assays because it did not affect the overall integrity of the cell, as growth phenotype and transcriptome profiles did not significantly deviate from the wild type properties. SmelC812, an antisense RNA of insertion sequence ISRm19, was postulated to prevent translation of its associated TRm19 transposase mRNA [21].

Induced overexpression of *ecpR1* led to abnormal cell elongation (Fig 2B). The mean cell length progressively increased after exposure to IPTG (Fig 2D). 30 hours post-induction 90% of the *ecpR1* overexpressing cells were abnormally long and 3% of the population additionally showed a branched morphology (sampling of 1000 cells). Similar abnormal cell morphologies have previously been reported in response to a variety of cell cycle perturbations that inhibit or overstimulate either DNA replication or cell division [32,34,39–41]. *ecpR1* overexpressing cells showed a ~2-fold decrease in generation time (~4 hours) compared to those overproducing the control sRNA (~2 hours), measured as the average time between two cell divisions monitored by time-lapse microscopy on TY rich medium (S3A Fig). Time-lapse microscopy also showed that after 30 hours of growth in presence of IPTG 38% of the elongated cells (n = 500) were not able to proceed to cell division and to resume growth after transfer to fresh medium lacking the inductor, compared to 4% of equally treated pSKControl$^+$ cells (S3B Fig). Furthermore, after three cycles of regrowing EcpR1 overproducing cultures on TY rich medium supplemented with IPTG, a 64% decrease in viable cells was observed (S3C Fig). Cells overproducing EcpR1 spread to a smaller halo (diameter 8 ± 2 mm) than the control (16 ± 1 mm) on soft agar (Fig 2C), but were still motile compared to a *visN* mutant incapable of swimming [42]. Finally, we checked alterations of the DNA content by fluorescence-activated cell sorting (FACS) analysis. 4 hours post-induction, cells with two genome copies started to accumulate in comparison to the control, and after 20 hours the majority of cells contained 2 or more genome equivalents (Fig 2E), further suggesting perturbations of the cell cycle.

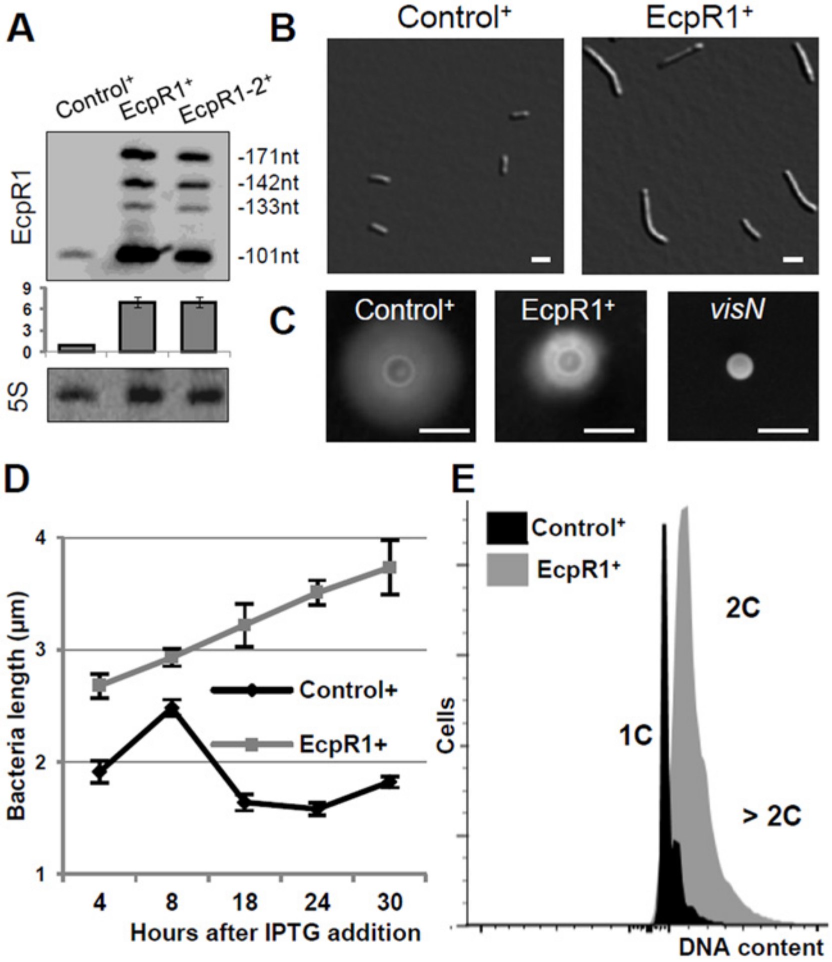


Fig 2. Elongated cell phenotype induced by *ecpR1* overexpression. (A) Northern blot detection of EcpR1 RNA variants in Rm4011 strains carrying either pSKControl$^+$ (Control$^+$), pSKEcpR1$^+$ (EcpR1$^+$), or pSKEcpR1-2$^+$ (EcpR1-2$^+$) 4 hours after induction with IPTG. Below, relative hybridization signals derived from the 101 nt EcpR1 species are plotted. The wild type level of EcpR1 in Control$^+$ cells (OD$_{600}$ of ~0.9) has been normalized to 1 (dashed line) and the sRNA levels in other conditions are correlated to that value. Mean results from three experiments are shown. Error bars indicate the standard deviation. **(B)** Cell morphology, **(C)** motility assay, **(D)** cell length, and **(E)** DNA content of *S*. *meliloti* strains overexpressing *ecpR1* or the SmelC812 control antisense RNA gene. The 2011*visN* mutant was used as negative control for swimming motility. 1C and 2C indicate one and two genome equivalents, respectively. Bars correspond to 2 µm in *B* and 5 mm in *C*. Error bars in *D* represent standard errors (n = 100 cells).

doi:10.1371/journal.pgen.1005153.g002

Because homologs of EcpR1 and cell cycle-related target candidates were also found in other members of the Rhizobiales, we asked whether overproduction of *S. meliloti* EcpR1 also leads to cell cycle defects in related species. This phenotype was conserved in the genera *Sinorhizobium* and *Rhizobium*, as IPTG-induced overexpression of *ecpR1* in *S. medicae*, *S. fredii*, *R. tropicii*, and *R. radiobacter* carrying plasmid pSKEcpR1$^+$ led to a similar proportion of elongated and branched cells as observed in *S. meliloti* (S4A Fig). In *R. etli* and *A. tumefaciens*, cell cycle associated defects were less abundant but FACS analysis confirmed an increased proportion of cells with more than two genome copies (S4B Fig).

Deletion of *ecpR1* attenuates competitiveness

The markerless 2011*ecpR1* mutant, missing the sequence of the full length 171 nt *ecpR1* variant, did not show distinct phenotypes in that it grew similarly to the wild type, even under the stress conditions which stimulated *ecpR1* expression (S5A–S5D Fig). After growth in rich medium or defined nutrient-limited minimal media until late stationary phase or after application of stress conditions growth recovery and cell viability (CFU/ml) were also not significantly affected compared to the wild type. Furthermore, the *ecpR1* deletion mutant was not impaired in symbiosis with its host plant *M. sativa* (S5E–S5H Fig).

The strong conservation and microsynteny suggests an evolutionary advantage conferred by the *ecpR1* locus. To support this hypothesis we determined whether the Rm2011 wild type has a fitness advantage over the *ecpR1* mutant. For this competitive growth assay, strains were labeled by a stable genomic integration of plasmids carrying either *egfp* or *mcherry* driven by a constitutive promoter. MOPS or Nutrient-limiting MOPS (MOPSlim) minimal media were inoculated with 2011 *mCherry* cells and either 2011 *egfp* or 2011*ecpR1 egfp* cells in a ratio of 1:1. eGFP:mCherry fluorescence ratios of the mixed cultures were measured and microscopy images were taken to determine the percentage of *egfp*-labeled bacteria (Fig 3, S6 Fig). After 7 days of cultivation, the 1:1 ratio was maintained indicating that all strains grew similarly, as we have previously observed when single-strain liquid cultures were grown in these conditions (S5 Fig). However, after the 7 days-old mixed cultures were diluted in fresh media, the proportion of the 2011*ecpR1 egfp* strain progressively decreased in the MOPSlim medium (S6C and S6D Fig). The mixture of 2011*egfp* and 2011*mCherry* cultures further on maintained the ~1:1 ratio, confirming that the fluorescence markers are neutral in the conditions tested (Fig 3). After three consecutive sub-cultivations, the *ecpR1* mutant only reached ~40% and ~20% of the population in MOPS and MOPSlim media, respectively (Fig 3). This implies a disadvantage of the *ecpR1* deletion mutant in recovery from late stationary cultures as compared to the wild type, particularly under nutrient limitation.

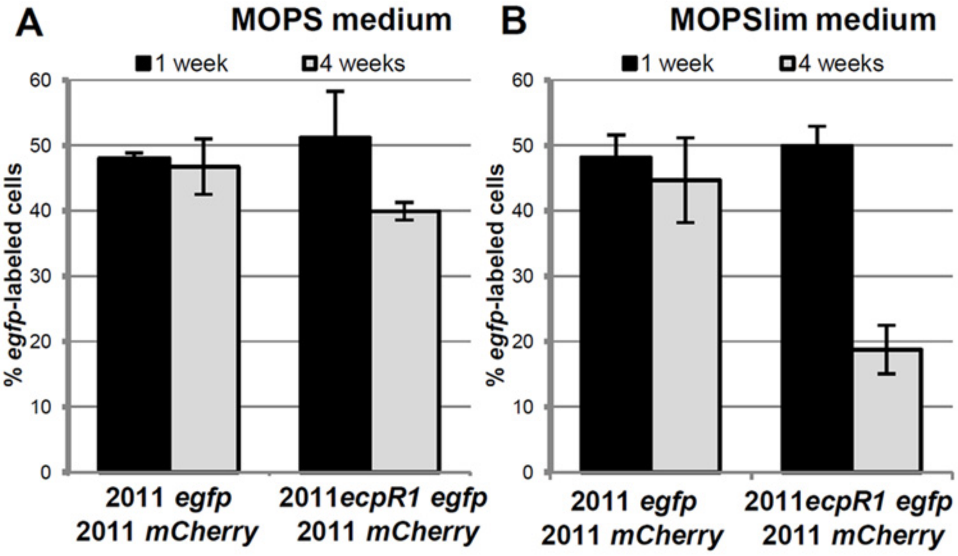


Fig 3. Lack of *ecpR1* reduces competitiveness of Rm2011. Mean percentage of *egfp*-labeled cells 1 and 4 weeks after mixing 2011*mCherry* with either 2011*egfp* or 2011*ecpR1 egfp* cells at a 1:1 ratio in MOPS (A) or MOPSlim media (B). Every week the mixed population was diluted 1000-fold in fresh media. The percentage of *egfp*-labeled cells was determined by microscopy. Error bars indicate the standard deviation of 3 biological replicates.

doi:10.1371/journal.pgen.1005153.g003

ecpR1 overexpression or deletion alters expression of genes related to cell cycle regulation

To obtain further clues to putative target genes of EcpR1 the cellular responses of the *S. meliloti* EcpR1 overproducing strain and the *ecpR1* deletion mutant (2011*ecpR1*) were characterized by microarray-based transcriptome profiling.

Differential gene expression upon EcpR1 overproduction: Genes displaying differential expression 15 minutes, 1 hour, and 4 hours post-induction of *ecpR1* overexpression in TY medium are listed in S2–S6 Tables. Only reporter oligonucleotides associated to the open reading frame or UTRs of 6 (15 minutes post-induction) and 20 (1 hour post-induction) protein-coding genes indicated transcript levels at least 1.6-fold lower than in the control. No genes were found to be upregulated after 15 minutes (except for *ecpR1* that was overexpressed) whereas RNA levels associated to 35 coding regions or UTRs including a number of ribosomal genes were upregulated after 1 hour. 4 hours post-induction, which corresponds to completion of one cell cycle in EcpR1 overproducing cells, transcript levels of 77 protein-coding genes were found to be changed (25 increased and 51 decreased). Several downregulated genes were related to cell cycle regulation and motility, which is in accordance with the observed phenotypes (Fig 2B–2E). Among these were *divJ* as well as the SMc00887-SMc00888 operon of unknown function that shares similarities with the *pleD-divK* operon located upstream of the *ecpR1* gene (Table 1). Previously, a decrease in SMc00887 and SMc00888 transcript levels was also found to be caused by mutation of *podJ* encoding a polarity factor [43]. The putative cell cycle-related SMc00888 gene was among the predicted EcpR1 targets (Table 1, position 22). Our transcriptome study also indicated lower representation of the *gcrA* 5'-UTR and increased levels of the long putative *dnaA* 5'-UTR region upstream of the predicted EcpR1 binding sites (Table 1 and Fig 4, vertical arrows), both among the top three ranked candidates of the computational EcpR1 target predictions (S1 Table). Most of the genes strongly differentially expressed upon EcpR1 overproduction are related to metabolism. We also found reduced levels of the 5'-UTR sequence of the ribonuclease gene *rne* 1 hour (M = -0.77) and 4 hours (M = -1.78) after

Table 1. qRT-PCR based verification of putative EcpR1 target genes displaying changes in transcript levels upon overproduction of EcpR1 as detected by global transcriptome profiling.

	Ratio of transcript levels: EcpR1 vs. SmelC812 overproduction		
Gene	**Description***	**Log_2 ratio (qRT-PCR)**	**M value (microarray)**
5'-UTR *gcrA* (-61 to -20)	cell cycle regulator GcrA	-1.03 ± 0.04	-0.76 ± 0.37
gcrA	cell cycle regulator GcrA	-2.06 ± 0.10	-0.41 ± 0.27
5'-UTR *dnaA*_5561 (-372 to -319)	chromosomal replication initiator DnaA	+1.23 ± 0.07	1.42 ± 0.40
5'-UTR *dnaA*_5562 (-222 to -174)	chromosomal replication initiator DnaA	+1.06 ± 0.04	0.74 ± 0.41
dnaA	chromosomal replication initiator DnaA	-1.54± 0.05	-
ctrA	cell cycle transcriptional regulator CtrA	-1.29 ± 0.04	-0.48 ± 0.32
divJ	sensor histidine kinase DivJ	-1.76 ± 0.10	-0.71 ±0.39
5'-UTR SMc00888 (-236 to -188)	2-component receiver domain protein SMc00888	-4.83 ± 0.17	-2.15 ± 0.63
SMc00888	2-component receiver domain protein SMc00888	-5.11 ± 0.24	-0.99 ±0.34
ftsZ1	cell division protein FtsZ1	-1.33 ± 0.05	-0.46 ± 0.42
pleC	sensor histidine kinase PleC, DivK phosphatase	-2.25 ± 0.07	-
minD	putative cell division inhibitor MinD	-0.46 ± 0.01	-

Log_2 change in transcript amount normalized to levels of the SMc01852 mRNA. Errors represent the standard deviation of three replicates. Positions of microarray reporter oligonucleotides relative to the start codon are given in brackets for 5'-UTR regions.

*Description of gene product or associated gene product.

doi:10.1371/journal.pgen.1005153.t001

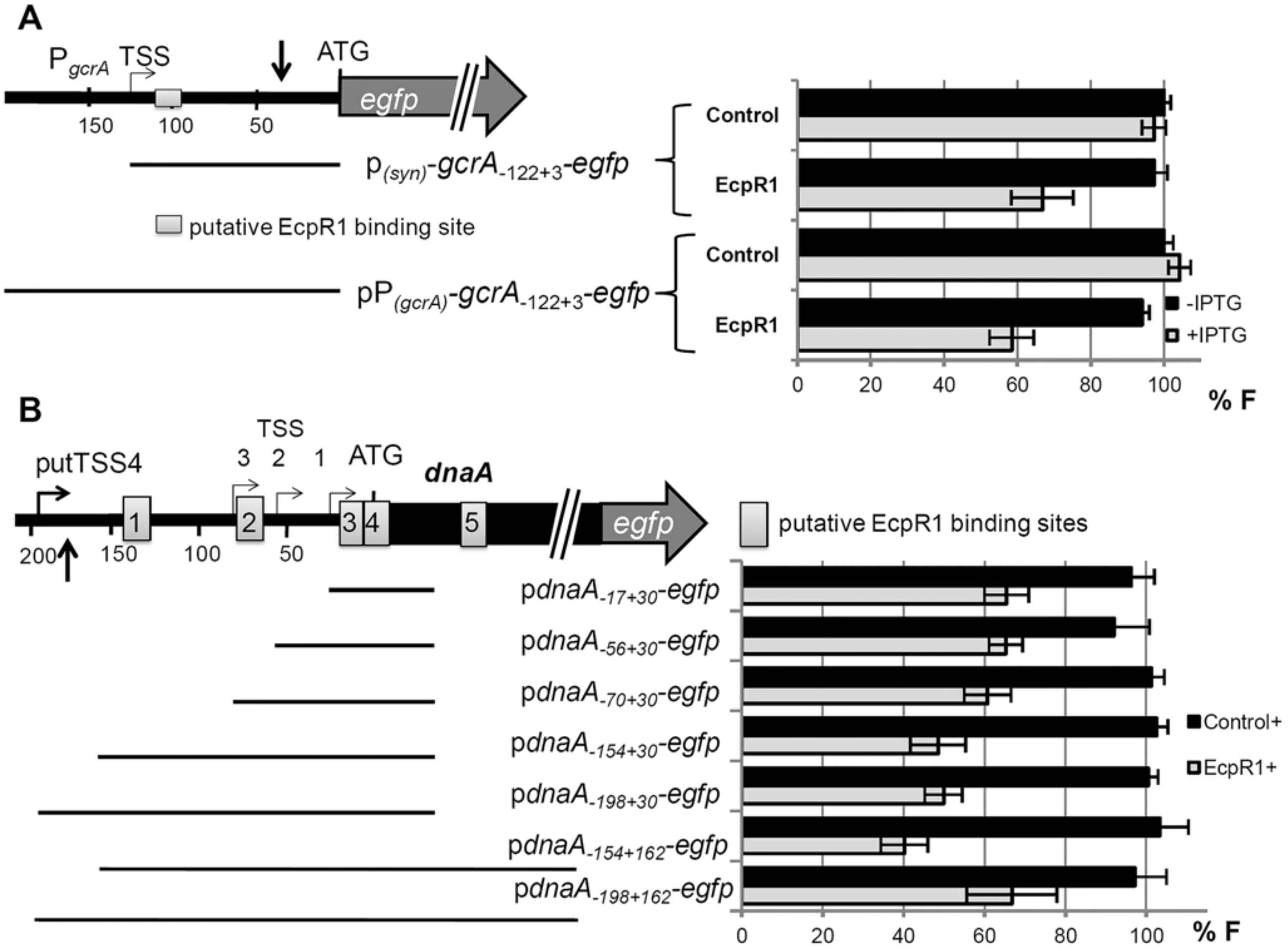


Fig 4. EcpR1 post-transcriptionally represses *gcrA* (A) and *dnaA* (B). Schematic representations of the genomic regions and the fragments (indicated by bars) translationally fused to *egfp*. Positions are denoted relative to the AUG; A is +1. Grey boxes indicate potential EcpR1-binding sites. Vertical arrows mark the regions covered by the oligonucleotide probes displaying altered signal intensities in the microarray hybridizations after *ecpR1* overexpression (see details in text). Means of relative fluorescence intensity values of Rm4011*ecpR1* co-transformed with the *ecpR1* or control SmelC812 overexpression plasmid, and the indicated reporter plasmid are shown below. The standard deviation represents at least three independent determinations of three double transconjugants grown in six independent cultures. Specific activities were normalized to the levels of the strain carrying the vector with the control RNA gene without IPTG added to yield percent relative fluorescence (% F).

doi:10.1371/journal.pgen.1005153.g004

ecpR1 overexpression. Moreover, the 5'-UTR sequence of *xerC* (M = +2.40), probably involved in chromosome segregation, and *mepA* (M = +1.15) encoding a homolog of peptidoglycan hydrolases, stood out among the upregulated transcripts 4 hours post-induction. qRT-PCR confirmed the observed changes in transcript levels of *dnaA*, *gcrA*, *divJ*, and SMc00888 in response to EcpR1 overproduction. Although not detected as differentially expressed in the microarray hybridizations, qRT-PCR showed reduced levels of the *ctrA*, *ftsZ1*, *pleC*, and *minD* transcripts in EcpR1 overproducing cells (Table 1).

Differential gene expression in the *ecpR1* deletion mutant compared to the wild type: The transcriptomes of Rm2011 and Rm2011*ecpR1* cells were compared during stationary growth in MOPS and MOPSlim media (S7–S10 Tables) since *ecpR1* expression is stimulated in the wild type under these conditions (Fig 1C, right panel). Reporter oligonucleotides associated to the open reading frame or UTRs of 18 (MOPS medium) and 17 (MOPSlims medium) protein-

Table 2. qRT-PCR based verification of putative EcpR1 target genes displaying expression changes in 2011*ecpR1* vs. Rm2011 wild type growing in MOPS or MOPSlim media.

		Ratio of transcript levels: EcpR1 vs. SmelC812 overproduction	
Gene	**Description***	**Log$_2$ ratio (qRT-PCR)**	**M value (microarray)**
5'-UTR *gcrA* (-61 to -20)	cell cycle regulator GcrA	0.81 ± 0.07 (MOPS)	-
		0.62 ± 0.06 (MOPSlim)	0.50 ± 0.28 (MOPSlim)
gcrA	cell cycle regulator GcrA	0.70 ± 0.10 (MOPS)	-
		1.31 ± 0.21 (MOPSlim)	-
dnaA	chromosomal replication initiator DnaA	0.80 ± 0.08 (MOPS)	0.67 ± 0.49 (MOPS)
		1.64 ± 0.33 (MOPSlim)	-
pleC	sensor histidine kinase PleC, DivK phosphatase	0.77 ± 0.07 (MOPS)	0.75 ± 0.09 (MOPS)
		1.28 ± 0.13 (MOPSlim)	1.02 ± 0.75 (MOPSlim)

Log$_2$ change in transcript amount normalized to levels of the SMc01852 mRNA. Errors represent the standard deviation of three replicates. Positions of microarray reporter oligonucleotides relative to the start codon are given in brackets for 5'-UTR regions.

*Description of gene product or associated gene product.

doi:10.1371/journal.pgen.1005153.t002

coding genes indicated transcript levels at least 1.6-fold lower than in the wild type control. Among them were reporters for the *dnaA* 5'-UTR region (positions -158 to -121 in MOPS and -222 to -174 in MOPSlim media), and *mepA*, both upregulated 4 hours after *ecpR1* overexpression.

In contrast, transcript levels of 27 (MOPS medium) and 44 (MOPSlims medium) protein-coding genes were found to be upregulated. Under both conditions *pleC*, ranking in the 5th position of the computationally predicted EcpR1 targets (S1 Table), displayed significantly higher transcript levels and was downregulated in EcpR1 overproducing cells (Table 1). Although *gcrA*, *dnaA*, and *pleC* microarray reporter signals did not pass all criteria set for the identification of differentially expressed genes, qRT-PCR indicated higher transcript levels of these genes in 2011*ecpR1* compared to the wild type (Table 2). This is in agreement with downregulation of these cell cycle-related genes upon *ecpR1* overexpression (Table 1).

In MOPSlim medium, several upregulated genes were related to cell division and cell wall degradation. Among those involved in cell division we found the *mraZ*-*mraW* genes (M = +1.30; +1.34) forming an operon with *ftsI*. The first gene of the *dll*-*ftsQ*-*ftsA* operon upstream of *ftsZ* (*dll*; M = +1.16) and *mltB2* (M = +1.18), both encoding homologs of peptidoglycan hydrolases, also appeared among the upregulated genes. Interestingly, several differentially expressed genes in the 2011*ecpR1* mutant harbour CtrA binding sites upstream the coding region, like *pleC*, *mraZ*, *mltB2*, and the genes coding for the PilZ-like protein SMc00999, the adenosylhomocystein hydrolase SMc02755, the putative transcriptional regulator SMc01842 and the hypothetical protein SMc03149. Beside this, in both media most of the differentially expressed genes with known functions were also related to metabolism. Among the strongly upregulated genes were the SMb20155-8 operon encoding the components of an ABC transporter (M = +2.57 to +3.22) and SMc03253 coding for an L-proline hydroxylase (M = +2.31). The latter was downregulated 15 min and 1 hour after induction of EcpR1 overproduction (M = -2.98 and -0.84, respectively).

However, looking for an overlap between the top target mRNA predictions ($P<0.005$) (S1 Table) and genes differentially expressed in the *ecpR1* overexpression or deletion strain (S2–S10 Tables) only genes related to cell cycle were identified.

EcpR1 post-transcriptionally represses the cell cycle master regulatory genes *gcrA* and *dnaA*

For experimental investigations, we restricted the set of EcpR1 target candidates to genes that fulfilled the following two criteria: (i) prediction by CopraRNA in the *Rhizobiaceae* with $P<0.005$ and (ii) decrease in transcript abundance upon *ecpR1* overexpression. These included *gcrA*, *dnaA*, *pleC*, *ftsZ*, *ctrA*, *minD*, and SMc00888. To this set we added *divK*, situated in the vicinity of the *ecpR1* locus (Fig 1A), and *divJ*. The corresponding mRNA sequences contain putative thermodynamically favored antisense interactions regions (S8 Fig).

To validate target mRNA candidates of EcpR1 *in vivo*, a double plasmid reporter assay was employed [44]. Target fragments comprising the native 5'-UTR [22] extended by the start codon or by a short 5'-part of the coding region were translationally fused to *egfp* in plasmid pR_EGFP and placed under the control of the constitutive synthetic P_{Syn} promoter [45]. All selected fragments contained the predicted EcpR1 interaction sequences. These plasmids were applied as reporter constructs to determine the post-transcriptional effect of induced EcpR1 overproduction on target mRNAs, while overexpression of the antisense RNA gene SmelC812 was used as control. This approach revealed EcpR1-induced down-regulation of reporter constructs corresponding to the top ranked predicted targets *gcrA* and *dnaA* ($P<0.0001$, Fig 4) but did not confirm the predicted regulatory effect of EcpR1 on the other cell cycle related target candidates (S7 Fig and S8 Fig). Since fluorescence mediated by the pSMc00888$_{-235+57}$-*egfp* reporter construct did not exceed the background level derived from the empty vector, we were unable to test this gene for EcpR1-induced regulation.

The EcpR1 binding region within the *gcrA* mRNA is located 13 nt downstream the TSS (position -122 relative to the AUG) (Fig 4A). The regulatory effect of EcpR1 on *gcrA* was assessed applying two different reporter constructs comprising the complete 5'-UTR fused to *egfp* either under the control of the constitutive P_{Syn} (plasmid p*gcrA*$_{-122+3}$- *egfp*) or the native *gcrA* promoter (plasmid pP*gcrA*$_{-122+3}$-*egfp*) (Fig 4A). Compared to the control, induced overexpression of *ecpR1* reduced p*gcrA*$_{-122+3}$-*egfp* and pP*gcrA*$_{-122+3}$-*egfp* mediated fluorescence to 34% and 42%, respectively (Fig 4A). Furthermore, activity of a chromosomally integrated *gcrA* 3'-*egfp* translational fusion [46] was reduced to 75% in response to *ecpR1* overexpression, validating the two-plasmid assay and confirming that posttranscriptional repression of EcpR1 results in reduction of GcrA protein level.

Five putative EcpR1 binding sites were identified within the *dnaA* mRNA (S1C Fig). Since different alternative ATG start codons have been assigned to *dnaA* in various rhizobial genomes, we affirmed the annotated ATG as translational start of *dnaA* in the Rm1021 genome [47]. None of the alternative start codons were functional when translationally fused to *egfp*. To test the post-transcriptional effect of EcpR1 overproduction on *dnaA* expression various fragments including all predicted binding sites or different subsets were translationally fused to *egfp* under the control of the constitutive P_{Syn} promoter (Fig 4B). Compared to the control, EcpR1 overproduction resulted in decreased activity of all reporter constructs, even when the shortest fragment was tested that only included the putative binding sites 3 and 4, overlapping the RBS and the start codon (Fig 4B).

A conserved GC-rich loop motif is essential for the regulatory function of EcpR1

Typically, sRNA sequences involved in mRNA base pairing are highly conserved, especially when binding multiple targets [48]. EcpR1 is predicted to fold into a secondary structure consisting of two hairpins (Fig 1A, S1A Fig): the 5' SL1 domain has a structurally conserved stem loop and a strongly conserved GC-rich loop motif (UCCGCCGCAUCU), which is predicted to be unpaired, while the SL2 domain includes a highly variable stem and a loop that contains the less conserved motif UCCUCG [27]. The predicted interaction region of EcpR1 mapped to the strongly conserved loop motif of SL1 which is part of the prevalent processed 101 nt transcript (Fig 1A, S1A Fig). Overproduction of an EcpR1 version starting from its second 5'-end (EcpR1$_{5'2}$) caused accumulation of this 101 nt core variant and the 142 nt version including the transcription termination sequence, and resulted in cell elongation (S9A and S9C Fig). This indicates that the 29 nt 5'-sequence of the full-length version is not required for provoking this phenotype.

Furthermore, overexpression of *ecpR1-2*, a full-length mutant variant carrying changes in 2 nt in the first loop sequence SL1, did not cause the alterations in cell morphology and DNA content previously observed upon overproduction of EcpR1 (Figs 2B–2D and 5A–5E). As EcpR1-2 conserved the predicted secondary structure of EcpR1 and Northern hybridizations confirmed the same level of overproduction of the mutant and the wild type variant (Fig 2A), we exclude that instability of the mutant RNA was responsible for the regulatory deficiency of EcpR1-2. This implies that the GC-rich loop motif is responsible for the cell cycle progression defects observed upon *ecpR1* overexpression. The single substitution G_{23} to C_{23} in EcpR1 (EcpR1-1) was not sufficient to destroy the regulatory activity of this sRNA (S9D Fig).

Moreover, overexpression of *ecpR1-2* did not post-transcriptionally repress *gcrA* and *dnaA* in the same strain background and culture conditions previously applied for EcpR1 (Fig 5E). Concordantly, 2 nt changes in the predicted target region within the *gcrA* 5'-UTR of the reporter fusion construct pP*gcrA*$_{-122+3}$-*egfp*, leading to construct pP*gcrA*$_{-122+3}$-BS-*egfp*, abolished fluorescence diminution caused by EcpR1 overproduction (Fig 5C and 5E). Introduction of 3 to 5 nt changes into the predicted binding sites 3, 4, or 5 within the *dnaA* mRNA only slightly mitigated the *ecpR1* overexpression-induced repression of reporter construct activities (Fig 5F). In the reporter constructs, substitutions in binding sites 3 and 4 (S1C Fig) were designed to avoid severe effects on translation of the mRNA because these binding sites overlapped the RBS and the start codon. Combined mutations of binding sites 3 and 4 abolished the negative regulatory effect of EcpR1 overproduction on the reporter construct activity (Fig 5F). This implies that the predicted interaction sites 1, 2 and 5 are not required for EcpR1-mediated repression of *dnaA* under the conditions tested.

Combination of the changes in the EcpR1 binding sites within the *gcrA* or *dnaA* 5'-UTRs and EcpR1-2 carrying the compensatory changes in the proposed interaction region partially restored the regulatory function of EcpR1-2. This further confirms the identified interaction regions in sRNA and mRNA (Fig 5B–5E). However, changing CCG to AAT in loop 1 of EcpR1 (EcpR1-3) destroyed its regulatory activity as expected, but the compensatory changes of GGC to TTA in the *gcrA* 5'-UTR did not restore it (S9F Fig). Northern blots showed that levels of EcpR1-3^{+} and EcpR1^{+} are similar (S9A Fig). Lack of restored regulation by compensatory mutations has already been reported for other sRNA-mRNA pairs [49–52] implying that both sequence and structure of the two RNAs are important for their interactions. The changes introduced affect not only the E score of the interaction, which dropped from -19.1 to -14.1, but also the nature of EcpR1 pairing at this position, which probably constitutes the sRNA seed region. This suggests that the binding strength mediated by the GC-rich sequence composition is important for the sRNA-mRNA interaction. Altogether, these data validate *gcrA* and *dnaA* as targets of EcpR1 and strongly suggests that this regulation is mediated through base pairing of the conserved GC-rich, single stranded region of EcpR1 with complementary GC-rich sequences of the target mRNAs.

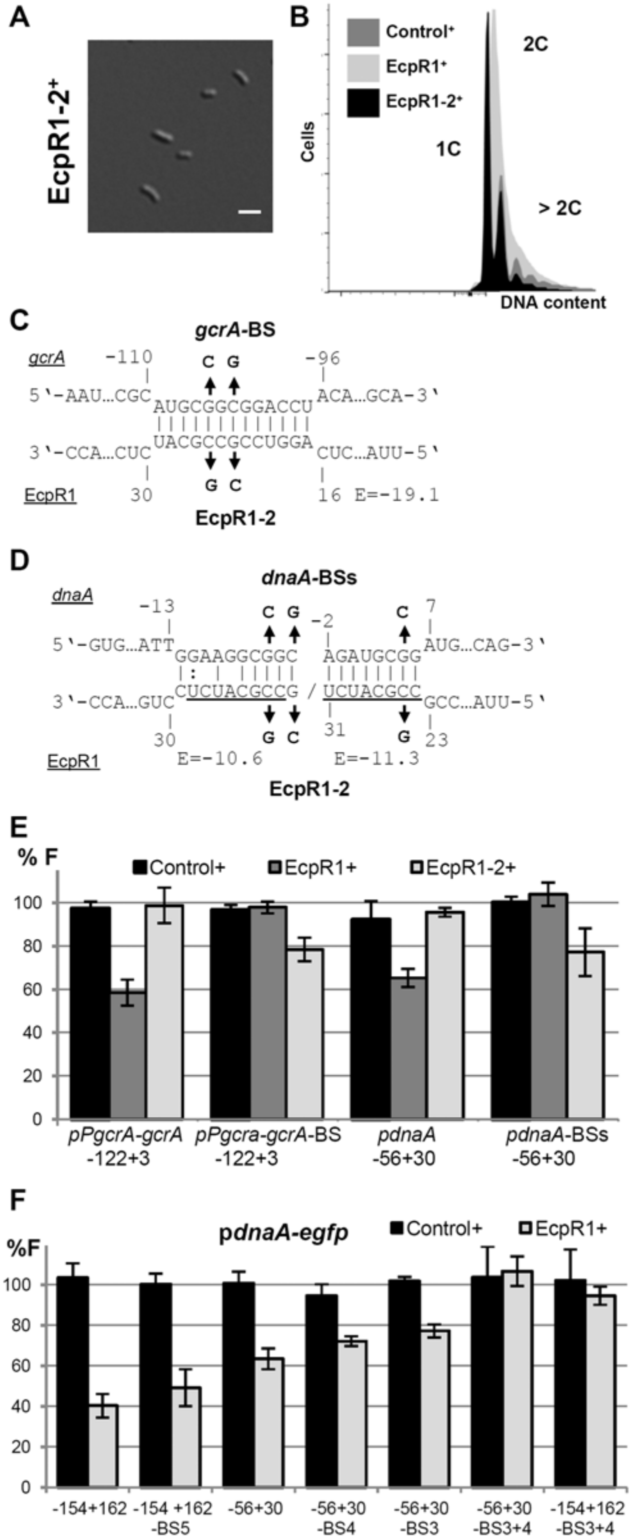


Fig 5. Validation of the predicted EcpR1 binding sites in the *gcrA* and *dnaA* mRNAs. Morphological phenotype **(A)** and DNA content **(B)** of Rm4011*ecpR1* overexpressing *ecpR1-2* carrying 2 nt exchanges in the predicted interaction region. The bar represents 2 μm. **(C, D)** Predicted duplexes between EcpR1 and either *gcrA* or *dnaA* mRNAs. Numbers denote positions relative to the AUG start codon of the mRNA and the second 5'-end of EcpR1. The predicted energy score (E) is indicated in kcal/mol. The nucleotide exchanges in the

mRNAs of *gcrA* (*gcrA*-BS-*egfp*) and *dnaA* (p*dnaA*-BSs-*egfp*) as well as in EcpR1 (EcpR1-2) are indicated in bold. **(E, F)** Fluorescence measurements of 4011*ecpR1* co-transformed with *ecpR1*, *ecpR1-2*, or control SmelC812 overexpression plasmids and the indicated reporter plasmids. Reporter constructs carried either native mRNA sequences derived from *gcrA* or *dnaA* or variants with mutations in predicted EcpR1 binding sites (BS). Fragments are delineated in Fig 4. Reporter construct activities were determined as in Fig 4.

doi:10.1371/journal.pgen.1005153.g005

EcpR1 function is Hfq-independent and requires RNase E to fully regulate *dnaA*

To further characterize the functional mechanism of EcpR1-dependent post-transcriptional regulation we tested the involvement of the RNA chaperone Hfq and the ribonuclease RNase E in this regulatory mechanism. Hfq is an RNA binding protein that canonically facilitates direct interaction of sRNAs and their mRNA targets and protects them from degradation in the absence of base pairing [53,54]. However, co-immunoprecipitation with epitope-tagged Hfq only detected 14% of the *S. meliloti trans*-sRNAs, excluding EcpR1, in cells grown under different stress conditions [55]. Accordingly, absence of Hfq did not compromise EcpR1 stability even 45 min after transcriptional arrest with rifampicin (Rf) as suggested by detection of similar levels of EcpR1 by Northern quantification in the Rm2011 wild type strain and the 2011*hfq* mutant (Fig 6A). In *S. meliloti*, knockout of *hfq* compromises growth, metabolism, motility, and stress adaptation in free-living bacteria [56,57]. In our study, microscopy analyses further showed abnormalities in cell morphology with some cells being filamentous and branched. Nevertheless, overexpression of *ecpR1* in 2011*hfq* caused cell elongation that was more severe than in the control (Fig 6B), suggesting that binding to Hfq is not required for EcpR1-mediated regulation of target mRNAs.

sRNAs associate with the C-terminal scaffold region of RNase E and other ribonucleases forming the so-called degradosome, which is recruited through base-pairing to the target mRNA to mediate its cleavage [58]. While the N-terminal catalytic domain of *E. coli* RNase E is essential for growth, the C-terminal region is dispensable and its deletion allows for testing the requirement of RNase E in sRNA-induced target mRNA degradation [58]. In *S. meliloti*, the C-terminal domain of RNase E is also non-essential, as either a mini-Tn*5* transposon insertion or a plasmid integration into codon 675 of *rne* led to viable cells, though moderately impaired in growth [59]. The 2011*rne*::Tn*5* mutant showed wild type morphology and displayed an elongated phenotype upon overexpression of *ecpR1* (Fig 6B). The same observation was made when comparing EcpR1 overproduction in strain 4011*ecpR1* versus 4011*ecpR1 rne675*. To further investigate whether this endoribonuclease is involved in EcpR1-mediated post-transcriptional regulation, the full-length reporter constructs pP*gcrA*$_{122+3}$-*egfp* and p*dnaA*$_{154+162}$-*egfp* were introduced to 4011*ecpR1 rne675* containing a plasmid either driving overproduction of EcpR1 or the control RNA SmelC812. A ~20% decrease in *gcrA* and *dnaA* reporter construct-mediated fluorescence was observed in Rm4011 *ecpR1 rne675* overexpressing *ecpR1* as compared to overproduction of SmelC812. In the 4011*ecpR1* strain carrying the complete *rne* gene the difference caused by EcpR1 overproduction was more pronounced for the *dnaA* reporter construct that showed a 39% lower reporter activity (Fig 6C). EcpR1-dependent decay of *gcrA* and *dnaA* mRNAs upon transcriptional arrest was assessed in 4011*ecpR1* either overexpressing *ecpR1* or the control RNA. Whereas decay of the *dnaA* mRNA was ~5-fold higher in the EcpR1 overproducing strain compared to the control strain after transcription inhibition, only a slight ~1.25-fold decrease in *gcrA* transcript levels was observed (Fig 6D). In summary, these data suggest that *dnaA* mRNA-EcpR1 interaction promotes RNase E-dependent mRNA degradation whereas EcpR1-mediated negative post-transcriptional regulation of *gcrA* is mostly independent of mRNA degradation and more likely due to translation inhibition of *gcrA*.

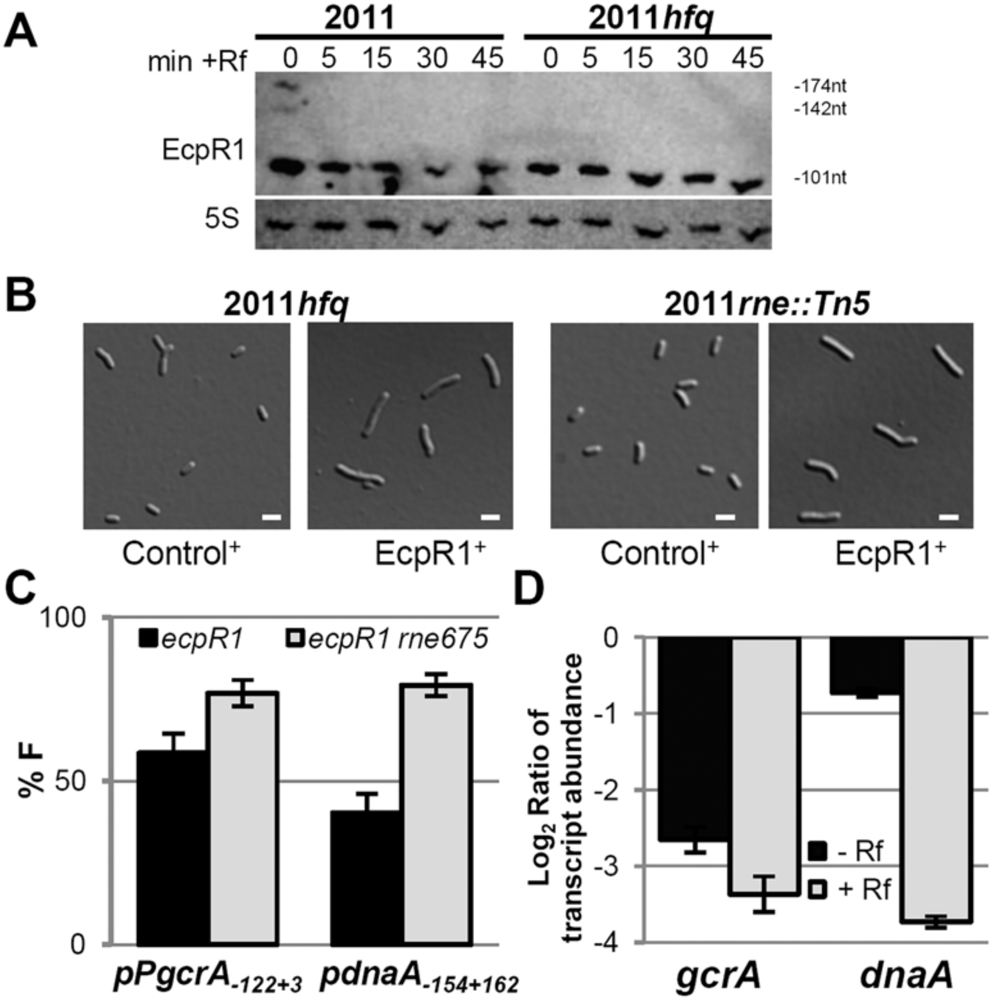


Fig 6. Hfq and RNase E activities are dispensable for EcpR1 overproduction-related cell elongation and post-transcriptional repression of *gcrA*. (A) Northern blot analysis of EcpR1 stability in Rm2011 and *hfq* mutant strains grown to early stationary phase (OD_{600} of 1.2, t = 0) and upon transcription arrest with Rf at indicated time points (in min). **(B)** Cell morphology of 2011*hfq* and 2011*rne*::*Tn5* mutants overexpressing either *ecpR1* ($EcpR1^+$) or the control RNA gene SmelC812 ($Control^+$) upon IPTG induction. Bars represent 2 μm. **(C)** Percentage of fluorescence in EcpR1 overproduction strains relative to the respective control strain overproducing SmelC812 in the Rm4011*ecpR1* or Rm4011*ecpR1 rne675* background co-transformed with plasmids carrying p*PgcrA-gcrA-egfp* or p*dnaA-154+162-egfp* translational fusions. **(D)** qRT-PCR analysis of *gcrA* and *dnaA* transcript abundance in Rm4011*ecpR1* $EcpR1^+$ after transcription arrest with Rifampicin for 5 minutes. Values were normalized to the SMc01852 transcript and the levels in the IPTG induced control strain overexpressing the SmelC812 RNA gene. Results from three independent experiments are shown. Error bars indicate the standard deviation.

doi:10.1371/journal.pgen.1005153.g006

Altered morphology caused by depletion of GcrA matches the elongated cell phenotype observed after EcpR1 overproduction

Recently, methylation-dependent binding of specific DNA motifs by orthologous GcrA proteins has been reported in several α-proteobacteria including *S. meliloti*, suggesting that this transcription regulator is functionally conserved in these bacteria [60]. Attempts to interrupt the *S. meliloti* 2011 *gcrA* coding region at the 98th codon by plasmid integration failed to produce any colonies. To further investigate the role of *gcrA*, a deletion mutant was constructed in

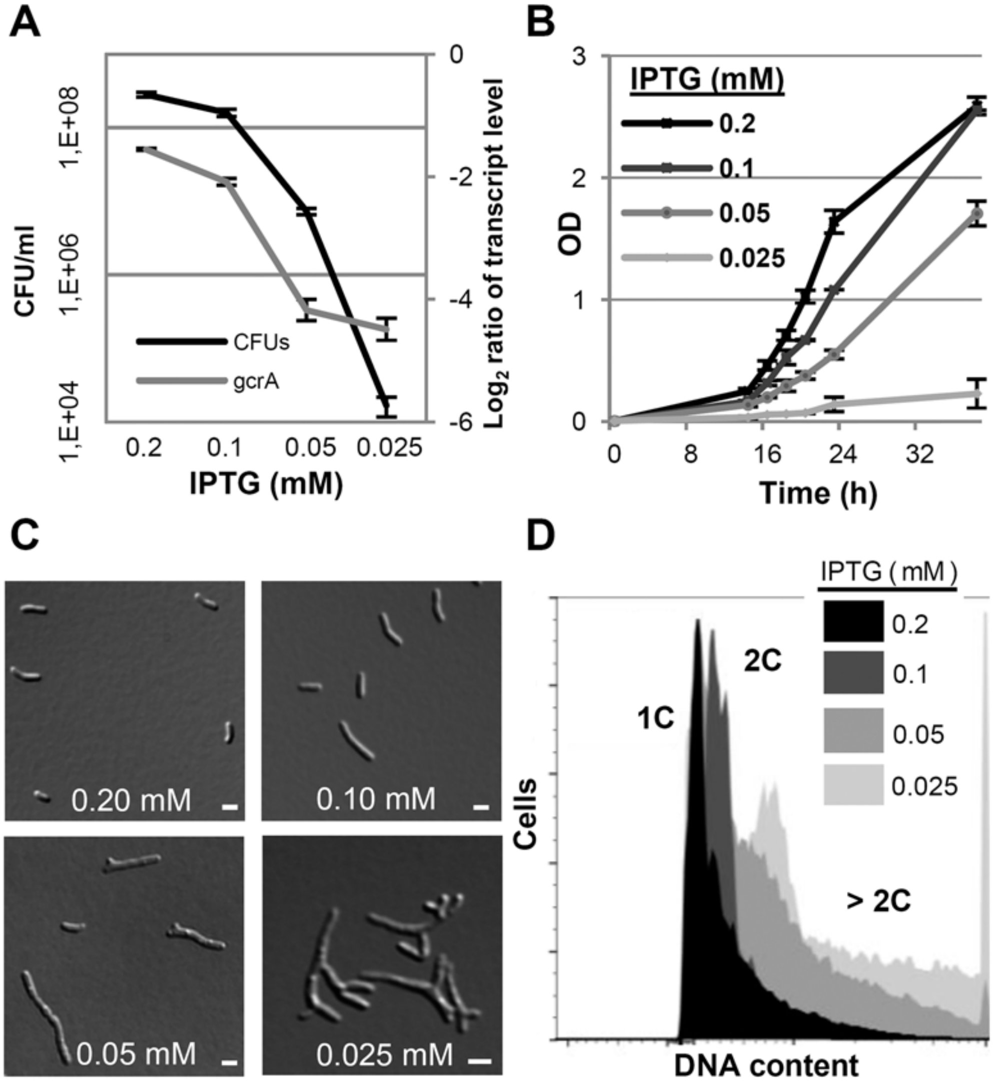


Fig 7. GcrA depletion phenotype in *S. meliloti*. qRT-PCR analysis of *gcrA* transcript abundance and colony forming units **(A)**, growth rate **(B)**, morphology phenotypes **(C)** and DNA content **(D)** of Rm2011*gcrA*-P$_{lac}$*gcrA* subjected to different IPTG concentrations for 16 hours. qRT-PCR values were normalized to the SMc01852 transcript and *gcrA* levels in overnight cultures of Rm2011. 1C and 2C indicate one and two genome equivalents, respectively. Error bars indicate the standard deviation. Bars denote 2 μm.

doi:10.1371/journal.pgen.1005153.g007

presence of a plasmid allowing for IPTG-induced expression of an ectopic copy of *gcrA* (2011*gcrA*-P$_{lac}$*gcrA*) since we also failed to obtain a *S. meliloti gcrA* deletion mutant. Strain 2011*gcrA*-P$_{lac}$*gcrA* was unable to divide in the absence of IPTG, suggesting that *gcrA* may be essential in *S. meliloti*.

To study the GcrA depletion phenotype, two independent clones of 2011*gcrA*-P$_{lac}$*gcrA* were grown in TY rich medium supplemented with 0.5 mM IPTG until early logarithmic phase. Cells were washed and subsequently cultured with different IPTG concentrations leading to lower *gcrA* transcript levels compared to the wild type (Fig 7A). Wild type-like growth was restored at ≥0.2 mM IPTG while lower concentrations hampered growth and cell viability (Fig 7A and 7B). The majority of 2011*gcrA*-P$_{lac}$*gcrA* cells grown with ≥0.2 mM IPTG displayed wild type-like morphology and harboured one or two genome equivalents. However, bacteria grown with 0.1 mM IPTG became elongated and the DNA content of the cells increased (Fig 7B and 7C). In contrast to the linear filamentous growth of a *C. crescentus* temperature sensitive *gcrA* mutant [60], 2011*gcrA*-P_{lac}*gcrA* cells cultured with ≤0.05 mM IPTG showed a tree-shaped morphology characterized by multiple branches (Fig 7C). Interestingly, the decrease in *gcrA* transcript level and the linear filamentous cell morphology observed in the mid-range of the tested IPTG concentrations resembled the phenotypic effects of induced EcpR1 overproduction.

Discussion

Microorganisms are often facing detrimental conditions unfavorable for cell proliferation such as biotic and abiotic stress factors or nutrient limitation. Therefore regulatory mechanisms adjusting replication initiation and cell cycle progression in response to environmental conditions are crucial for survival. Bacteria have evolved diverse mechanisms to couple perception of stress conditions to a cellular response that triggers a slow down or arrest of cell cycle progression [10]. The most prominent regulatory route for cell cycle control in response to nutrient deprivation involves the stringent response common to diverse bacteria. The stringent response second messenger ppGpp was shown to cause a G1 arrest in *E. coli*, *C. crescentus*, and *Bacillus subtilis* by modulating abundance or activity of proteins involved in DNA replication, such as DnaA or the primase DnaG. However, the underlying mechanisms are largely unknown. Recently, accumulation of unfolded proteins upon abiotic stress was reported to induce targeted degradation of DnaA resulting in cell cycle arrest in *C. crescentus* [61]. Inhibition of cell division mediated by the SOS response was observed in response to DNA damage gaining time for repair. Targeting of divisome components has been shown to be inherent to this DNA damage response in *E. coli* and *C. crescentus*.

In this study, we add *trans*-sRNA mediated regulation as another layer contributing to these diverse mechanisms linking stress factor sensing to the cell cycle engine. To the best of our knowledge, EcpR1 constitutes the first example of a *trans*-sRNA directly post-transcriptionally modulating expression of two cell cycle related genes in prokaryotes. Despite the effort invested in the model organism *C. crescentus* to identify sRNAs exhibiting cell cycle-dependent expression profiles [19], the connection between them and the cell cycle engine remained unproven.

To date, two antisense RNAs related to bacterial cell cycle genes have been identified: the defective prophage-encoded DicF RNA in *E. coli*, and asDnaA in *Salmonella enterica*. DicF inhibits translation of the cell-division protein FtsZ when overexpressed [62], while *asdnaA* is expressed in stationary phase and under other stress conditions and seems to increase stability of the *dnaA* mRNA by an unknown mechanism [63]. A few sRNAs have been reported to be involved in bacterial cell differentiation processes that may include modulation of cell cycle control. *trans*-sRNA Pxr negatively regulates fruiting body formation in *Myxococcus* [64]. In *Chlamydia*, the conserved IhtA sRNA translationally inhibits the histone-like protein Hc1 that is involved in compaction of the chromatin into metabolically inert forms during host infection [65,66]. In *E. coli*, the plasmid-encoded Rcd RNA indirectly regulates cell growth to ensure plasmid maintenance by binding to a protein involved in indole metabolism [67].

Quick responses to suddenly arising adverse conditions provide an adaptive advantage to the cell. Riboregulators have the potential to act faster as regulatory proteins since RNA is the first product of gene expression. The most prevalent mechanisms of *trans*-sRNA mediated riboregulation affect mRNA translation and stability, which also are most likely the modes of action of EcpR1 on the *gcrA* and *dnaA* target mRNAs in the α-proteobacterium *S. meliloti*. We speculate that affecting synthesis of cell cycle master regulators at this post-transcriptional level is an advantageous complementary mechanism to stress-stimulated proteolysis as reported for DnaA in the distantly related α-proteobacterium *C*. *crescentus* [61].

Most sRNAs are conserved only among closely related species, but EcpR1 shows a broad distribution within the Rhizobiales, including organisms with different lifestyles, such as pathogens (e.g. *Agrobacterium)* and diazotrophic plant endosymbionts. EcpR1 overproduction-induced perturbations of cell cycle progression in several species harboring members of the SmelC291 (EcpR1) RNA family also imply functional conservation of this sRNA. However, deletion of *ecpR1* did not cause significant differences in cell growth or viability, but attenuated competitiveness with the wild type. Since sRNAs primarily act to fine-tune stress responses that commonly rely on redundant bacterial pathways [8] sRNA mutants frequently do not show significant phenotypes under laboratory conditions.

The majority of the bacterial sRNAs characterized so far accumulate under stress conditions [68] as does EcpR1, suggesting that this sRNA likely constitutes an adaptive factor that contributes to prevent cell-cycle progression when cells must slow down proliferation. Tight control of EcpR1 levels are likely to be crucial since an excessive amount resulted in a considerable proportion of cells that were not able to resume growth after *ecpR1* overexpression had been stopped. This is in agreement with a more moderate induction of EcpR1 production under stress conditions in the native situation. We obtained evidence that transcription of *ecpR1* driven by an RpoD-type promoter is stimulated by ppGpp, placing EcpR1 in the stringent response regulon of *S*. *meliloti*. This finding is intriguing in light of the role of the stringent response in coupling nutrient status to cell cycle control.

Interestingly, the elongated phenotype of cells overexpressing *ecpR1* resembles that of differentiated nitrogen fixing bacteroids inside plant root nodules and recently, it has been found that nodule-specific cysteine-rich (NCR) peptides triggering rhizobial genome endoreduplication perturbed expression of *dnaA*, *gcrA*, and *ctrA* [69]. In our study, EcpR1 was not detected in *M*. *sativa* mature root nodules implying that *ecpR1* is not expressed in bacteroids. This is in agreement with a transcriptome study of individual zones of the root nodule which determined only low levels of EcpR1 in the symbiotic zone containing mature bacteroids and found the highest concentration of EcpR1 in the interzone where bacteroid differentiation occurs [70].

The confirmed target genes of EcpR1, *dnaA* and *gcrA*, encode key regulators of a complex regulatory circuit governing replication initiation and cell cycle progression. Despite subtle differences, the architecture of this regulatory circuit displays a high degree of similarity in *S*. *meliloti* and *C*. *crescentus* [25,26]. In *C*. *crescentus*, DnaA activates *gcrA* expression [13]. However, computational comparisons did not predict a significant DnaA binding motif in the promoter sequence of the *S*. *meliloti gcrA* gene, but upstream of *divJ* encoding a kinase/phosphatase involved in control of CtrA activity and upstream of SMc00888 encoding a DivK homolog [25,26]. GcrA controls expression of a multitude of target genes including *ctrA*. CtrA-binding motifs have been identified in the promoter regions of *pleC*, *minD*, SMc00888 and *fts*, and the *fla* genes [26]. In *S*. *meliloti*, transcriptome profiling and qRT-PCR assays suggest a direct or indirect effect of EcpR1 overproduction on a number of genes that are core components or known to be under control of this regulatory circuit, further supporting the modulating effect of EcpR1 in the regulatory context of cell cycle control. The enhanced levels of the *dnaA* 5’UTR caused by EcpR1 overproduction may be explained by mechanisms favoring accumulation of the 5’UTR (such as stabilization or attenuation) in conjunction with DnaA autoregulation as reported for *E*. *coli* [71] and feedback regulation increasing levels of DnaA in GcrA-depleted *C*. *crescentus* cells [72]. Such mechanisms may compensate for EcpR1-mediated negative post-transcriptional regulation of *dnaA*. Although significant, transcriptional changes of the cell cycle-related genes were rather low. In the non-synchronized cultures, this might have been due to heterogeneous expression of such genes dependent on the cell cycle state as has been described for *gcrA* in *S. meliloti* and other cell cycle-dependent genes whose transcription varies during cell cycle progression [25,46].

Computational target predictions for EcpR1 suggested several cell cycle related target mRNAs among the top 50 candidates ($P<0.005$), albeit transcriptome and *in vivo* interaction studies only provided evidence for a direct interaction with *gcrA* and *dnaA* mRNAs, ranking in positions 1 and 3, respectively. Still, we cannot exclude that further interactions occur which the two-plasmid assay failed to detect. Similarities between phenotypes caused by EcpR1 overproduction and modest GcrA depletion suggest that a decrease in GcrA concentration contributed to this perturbation of cell cycle progression. DnaA-depletion has been reported to go along with an increase in cell length, while DNA synthesis is arrested [16]. These elongated cells contained only one chromosome, in contrast to the *ecpR1* overexpressing cells that showed an increase in cell length and DNA content.

Although *gcrA* and *dnaA* promoter regions have been extensively studied in *C. crescentus* [13,72,73], the functions of the long 5'-UTRs are still unknown in both organisms. Here, we obtained evidence that these 5'-UTRs are involved in *trans*-sRNA mediated post-transcriptional regulation in *S. meliloti*. Our experiments indicate that degradation of the *gcrA* mRNA was not significantly promoted by *ecpR1* overexpression. Yet, the output of a reporter gene fused to the *gcrA* 5'-UTR was considerably reduced. This is indicative of EcpR1 rather affecting translational efficiency than stability of the *gcrA* mRNA. However, the single binding site for EcpR1 was identified close to the TSS far upstream of the RBS. sRNA-mediated translational control mostly involves its binding to sequences surrounding the RBS, preventing the ribosome from initiating translation. So far, alternative mechanisms of translational control have been poorly studied, but other models of sRNA repression, such as competing with a “RBS standby site” or pairing with a translation enhancer element have been proposed [74]. In contrast, stability of the *dnaA* mRNA was negatively affected by enhanced levels of EcpR1 and the regulatory effect of this sRNA was significantly alleviated by a C-terminal truncation of RNase E, suggesting that EcpR1 promotes *dnaA* mRNA degradation. Assuming that EcpR1-induced cell cycle perturbation is mainly due to translational inhibition of *gcrA*, these data are in agreement with maintaining this phenotype in the background of the RNase E truncation.

Computational analysis predicted five sequence motifs in the *dnaA* mRNA that are likely to form a stable duplex with EcpR1, which is an exceptionally high number for these types of interactions. Our data strongly suggests that the two binding sites overlapping the RBS and the start codon are sufficient for and synergistically enhance the regulatory effect of EcpR1 on the *dnaA* mRNA under the conditions tested. A conserved GC-rich sequence in loop 1 of EcpR1 was consistently found to be involved in the interactions with these two binding sites in the *dnaA* and one binding site in the *gcrA* mRNA. In bacteria and plants, multiple binding of a target mRNA by a *trans*-sRNA mediated by the same interaction region is a rare finding, although frequently observed for regulatory non-coding RNAs in animals. Binding of multiple target sequences in bacterial mRNAs has been reported, but usually involves different interaction regions of the sRNA. Examples are the MicF sRNA that binds to the *lpxR* mRNA both at the RBS and in the coding sequence [75], as well as the polycistronic mRNA *manXYZ* which is targeted at the RBS and in the intergenic region through overlapping interaction regions of the sRNA SgrS [76].

EcpR1 broadens the unprecedented discovery of prokaryotic sRNA functions of the last two decades. Although additional biological roles of EcpR1 remain to be investigated, stress-induced stimulation of EcpR1 production and its posttranscriptional effect on *gcrA* and *dnaA* suggest an additional level of regulation contributing to a rapid and robust response of the cell cycle machinery to adverse environmental conditions.

Materials and Methods

Bacterial strains, plasmids and growth conditions

Bacterial strains and plasmids are listed in S11 Table. *E. coli* strains were routinely grown at 37°C in LB medium and rhizobial strains at 30°C in complex tryptone yeast (TY) medium [77] or in modified MOPS-buffered minimal medium [78] (MOPS-MM: MOPS, 10 g l^{-1}; mannitol, 10 g l^{-1}; NH_4Cl, 1 g l^{-1}; NaCl, 0.1 g l^{-1}; $MgSO_4$, 0.246 g; $CaCl_2$, 250 mM; $FeCl_3{\bullet}6H_2O$, 10 mg l^{-1}; H_3BO_3, 3 mg l^{-1}; $MnSO_4{\bullet}4H_2O$, 2.23 mg l^{-1}; biotin, 1 mg l^{-1}; $ZnSO_4{\bullet}7H_2O$, 0.3 mg l^{-1}; $NaMoO_4{\bullet}2H_2O$, 0.12 mg l^{-1}; $CoCl_2{\bullet}6H_2O$, 0.065 mg l^{-1}, pH 7.2). Nutrient-limiting MOPS (MOPSlim) was modified as follows: mannitol, 2 g l^{-1}; NH_4Cl, 0.3 g l^{-1}; NaCl, 0.05 g l^{-1}; $MgSO_4$, 0.1 g l^{-1}. MOPS-C and -N lack mannitol or ammonium chloride, respectively. Antibiotics were added to solid media when required to the following final concentrations (mg/ml): streptomycin (Sm) 100 for *Rhizobium* and 600 for *Sinorhizobium* strains; nalidixic acid (Nx) 10; ampicillin (Ap) 200; tetracycline (Tc) 10; gentamycin (Gm) 40; rifampicin (Rf) 50; chloramphenicol 20; and kanamycin (Km) 50 for *E. coli* and *Rhizobium* and 180 for *Sinorhizobium* strains. For liquid cultures, the antibiotic concentration was reduced to 50%. IPTG was added to a final concentration of 0.5 mM to exponential phase cultures (OD_{600} of 0.3 to 0.4), unless other conditions are indicated. For stress induction, media of exponentially growing cultures were modified as described [22] and harvested 1 hour later. Motility assays, were carried out by dispensing 3 μl aliquots of the corresponding bacterial suspension (OD_{600} of 0.9 to 1) on soft agar plates and incubating at 30°C for 5 days. Plant nodulation assays were basically performed as described before [79].

RNA isolation and northern hybridization

RNA was isolated from bacterial cultures and from 28 days old *M. sativa* cv. Eugenia root nodules with the miRNeasy Mini Kit (Qiagen). Nodules covered with liquid nitrogen were ground to powder in a mortar before RNA isolation. For Northern blot detection of RNAs, 4 μg total RNA was separated on 10% polyacrylamide gels containing 7 M urea and transferred onto nylon membranes by semi-dry electroblotting. An EcpR1-specific DIG-labeled DNA probe was used for hybridization (50°C) and detection was performed using the DIG Luminescent Detection Kit (Roche) following the manufactures instructions. Size was determined in relation to an RNA molecular weight marker (NEB).

Construction of the *S. meliloti* mutants and derivative strains

GeneSOEing was used to construct the marker-free deletion of the chromosomal *ecpR1* locus and the strain with mutations in the *ecpR1* σ^{70}-dependent promoter -10 region using the internal complementary primers listed in S12 Table. The digested PCR fusion product containing *ecpR1* flanking sequences or the *ecpR1* locus region carrying changes in the promoter -10 region were cloned into suicide vector pK18mobsacB, respectively. Double cross-over events were selected as previously described [80] and checked for the targeted deletion by PCR, sequencing and Northern analyses. To create a conditional depletion mutant, the *gcrA* locus was also deleted by geneSOEing, but this deletion was introduced to *S. meliloti* harbouring plasmid pSRKGm containing the *gcrA* gene under control of the IPTG inducible P_{lac} promoter (P_{lac}*gcrA*). Double recombinants were selected on medium supplemented with IPTG and subsequently grown on agar with and without IPTG. Strains exhibiting IPTG-dependent growth were selected and the chromosomal *gcrA* deletion was checked by PCR amplification and sequencing of the *gcrA* chromosomal locus.

For IPTG induced overexpression of *ecpR1* an indirect *sinR*-*sinI* based system was applied. In *S. meliloti*, the LuxR-type transcription regulator SinR strongly activates the promoter of the N-acyl homoserine lactone synthase encoding gene *sinI* [81]. The complete sequence of the *sinR* gene and the *sinR*-*sinI* intergenic region containing the *sinI* promoter were fused to the TSS of the control sRNA gene SmelC812 or the corresponding 5'-end of *ecpR1* by geneSOEing. The resulting fragments were inserted into pSRKKm to generate the expression plasmids that were transferred by conjugation to Rm4011 (*expR*$^-$ *sinI*$^-$) to minimize background expression. A PCR-based mutation strategy was used to replace specific nucleotides within the corresponding plasmid constructs as described before [82] using the internal complementary primers listed in S12 Table.

eGFP-mediated fluorescence constructs and assays

For construction of *ecpR1* promoter-*egfp* fusions the corresponding genomic fragments (Fig 1B) were amplified and cloned into plasmid pPHUtrap, a derivative of pPHU231 [83] containing a promoterless *sinI* 5'-UTR fused to *egfp*. *S. meliloti* cells carrying the *ecpR1* promoter fusions were grown until stationary phase and 100 μl of the cultures were transferred to a 96 well microtiter plate and measured as described below. To accurately compare the activities of the promoter fusions at different OD_{600} values (0.6, 1.2, and 2.8), cells harvested at OD_{600} of 1.2 and 2.8 were diluted to OD_{600} of 0.6 before being transferred to the 96 well microtiter plate for measurement.

To determine EcpR1 target mRNA regulation *in vivo*, plasmid pR_EGFP [44] was used to constitutively express 5'-UTR translational fusions of the predicted target genes from its native TSS [22]. The reporter plasmids were transferred by conjugation to Rm4011*ecpR1* harboring plasmids pSKControl$^+$ or pSKEcpR1$^+$. Three double transconjugants for each RNA-target fusion combination were grown to mid-exponential phase (OD_{600} of 0.3 to 0.4) and 100 μl aliquots of IPTG treated and untreated cultures were transferred to a 96 well microtiter plate and incubated at 30°C with shaking for 8 hours.

OD_{600}, eGFP and mCherry-mediated fluorescence were measured in the Infinite M200 Pro microplate reader (Tecan). Fluorescence values were normalized to the culture OD_{600}, and background F/OD ratios from strains harboring the corresponding empty plasmid (pPHUtrap or pR_EGFP) were subtracted from those mediated by each reporter construct.

Competitive growth assay

For estimation of the relative fitness, Rm2011 and 2011*ecpR1* were labeled with *mCherry* or *egfp* by single integration of either plasmid pKOSm or pKOSe, both pK18mobII derivatives carrying P_{T5}:*mCherry* or P_{T5}:*egfp* cassettes follow by a T7 terminator site and a 800 bp fragment from *recG*. Strains were individually grown in MOPS or MOPSlim media starter cultures overnight and bacteria were then diluted in the same fresh media to OD_{600} of 0.005 and mixed at a ratio of 1:1 in a final volume of 30 ml. During a 4 weeks period, every seven days of incubation eGFP and mCherry fluorescence of the cultures were measured and the mixed population was diluted 1000-fold in fresh media. One and four weeks after the first mixed inoculation microscopy images were taken to determine the percentage of eGFP- and mCherry-labeled bacteria.

Microarray-based gene expression profiling

Four independent bacterial cultures of Rm4011 carrying pSKEcpR1$^+$ or pSKControl$^+$ or either Rm2011 or 2011*ecpR1* were grown in 100 ml of the corresponding medium for each experiment. Cells were harvested in the indicated conditions (15 minutes, 1 hour, and 4 hours after

IPTG induction or in the stationary phase of growth) and RNA was isolated. cDNA synthesis, Cy3- and Cy5 labeling, hybridization, image acquisition and data analysis were performed as previously described [84]. Normalization and t-statistics were carried out using the EMMA 2.8.2 microarray data analysis software [85]. Genes and 5'-/3'-UTRs with P-value$\leq$0.05 and M$\geq$0.7 or $\leq$−0.7 were included in the analysis. The M value represents the $\log_2$ ratio of both channels. Transcriptome data are available at ArrayExpress (Accession No. E-MTAB-3389).

Quantitative RT-PCR analysis

qRT-PCR was carried out in a qTOWER Thermal Cycler (Analytik Jena, Germany) using the KAPA SYBR FAST One-Step qRT-PCR Kit and 50 ng of RNA per reaction (5 μl). The ratios of transcript abundance were calculated as the $2^{-\Delta CT}$ mean average of 3 replicates, where CT indicates the level of gene expression in the specified strain relative to the expression in the control strain. The uniformly expressed gene SMc01852 [86] was used to normalize the gene expression data.

Microscopy

Bacteria were visually examined by differential interference contrast and epifluorescence or highly inclined laminated optical sheet microscopy (Tokunaga) using a Nikon Eclipse Ti-E equipped with 100x CFI Apo TIRF Oil objective (numerical aperture of 1.49) with AHF HC filter sets F36-513 DAPI (excitation band pass 387/11 nm, beam splitter 409 nm, emission band pass 447/60 nm), F36-504 TxRed (ex bp 562/40 nm, bs 593 nm, em bp 624/40 nm) and F36-525 eGFP (exc bp 472/30 nm, beam splitter 495 nm, em bp 520/35 nm). Living cells grown to the desired condition were directly placed on 1% TY agarose pads. Images were acquired with an Andor iXon3 885 EMCCD camera. Image acquisition, measurements and adjustment were done with Nikon NIS elements 4.0 software. For time-lapse analysis images were acquired every 15 minutes at 30°C.

Fluorescence-activated cell sorting (*FACS)*

To identify DNA content of single cells, 200 μl of culture grown to the desired condition was harvested and fixed in 70% cold ethanol. For examination, fixed cells were washed twice and resuspended in 200 μl of 50 mM sodium citrate buffer, and DNA was stained with 50 μg/ml Hoechst 33342. Acquisition was done on a BD Biosciences LSRII flow cytometer and analyzed using FlowJo 10 software. Each histogram represents the analysis of 50,000 cells.

Bioinformatics tools

sRNA secondary structures were predicted with RNAfold [36] and represented with VARNA [87]. The full-length EcpR1 sequence was scanned for antisense interactions within several genomes using CopraRNA with standard parameters [28]. *S. meliloti* (NC_003047) was included as organism of interest in all rounds of genome-wide target predictions, first together with seven closely related *Rhizobiaceae* species belonging to the genera *Sinorhizobium* (NC_009636, NC_012587), *Agrobacterium* (NC_011985, NC_003063, NC_011988), and *Rhizobium* (NC_007761 and NC_008380). The second group included NC_008254, NC_014923, and NC_002678 from the genus *Mesorhizobium (Phyllobacteriaceae)* and the third group representatives of the *Xanthobacteriaceae* belonging to the genera *Starkeya* (NC_014217), *Xanthobacter* (NC_009720), and *Azorhizobium* (NC_009937). Finally, predictions included the same *Xanthobacteriaceae* representatives together with *Methyocella* (NC_011666) and *Beijerinckia* (NC_010581) (*Beijerinckiaceaceae*), and *Rhodomicrobium* (*Hyphomicrobiaceae*). Predicted individual sRNA-mRNA duplexes were further confirmed with IntaRNA [88] and RNAup [36]. Functional enrichment of EcpR1 top target candidates was assessed applying Fisher's exact test. For this, the fisher.test function from R statistics [89] was employed with the "alternative" parameter set to "greater". Based on homology search, 53 *S. meliloti* genes are cell cycle related. Of these, 50 are present in the total CopraRNA prediction list (length = 4962) and seven of these 50 are in the top predicted target list (length = 89) at P< = 0.01. In R notation, this leads to the following matrix for fisher.test function: matrix(c(7,43,82,4830),nrow = 2, ncol = 2). The *S. meliloti ecpR1*–100 region was BLASTed with default parameters against all currently available bacterial genomes and several regions exhibiting significant similarities (80–100% similarity) were used to generate automated alignments.

Supporting Information

S1 Table. CopraRNA results of *S. meliloti* target candidates predicted for the SmelC291 family (EcpR1) of homologous sRNAs present in closely related *Rhizobiaceae* species belonging to the genera *Sinorhizobium*, *Agrobacterium*, and *Rhizobium*. Cell cycle related candidates used for the enrichment analysis are denoted in bold and experimentally confirmed targets are underlined.
(PDF)

S2 Table. Genes and 5'-/3'-UTRs differentially expressed 15 minutes after induction of EcpR1 overproduction (P-value≤0.05 and M≥0.7 or ≤−0.7). The M value represents the $\log_2$ ratio of transcript levels.
(PDF)

S3 Table. Genes and 5'-/3'-UTRs displaying decreased expression 1 hour after induction of EcpR1 overproduction (P-value≤0.05 and M≥0.7 or ≤−0.7). The M value represents the $\log_2$ ratio of transcript levels.
(PDF)

S4 Table. Genes and 5'-/3'-UTRs displaying increased expression 1 hour after induction of EcpR1 overproduction (P-value≤0.05 and M≥0.7 or ≤−0.7). The M value represents the $\log_2$ ratio of transcript levels.
(PDF)

S5 Table. Genes and 5'-/3'-UTRs displaying decreased expression 4 hours after induction of EcpR1 overproduction (P-value≤0.05 and M≥0.7 or ≤−0.7). The M value represents the $\log_2$ ratio of transcript levels. Cell cycle related candidates are indicated in bold and experimentally confirmed targets are underlined.
(PDF)

S6 Table. Genes and 5'-/3'-UTRs displaying increased expression 4 hours after induction of EcpR1 overproduction (P-value≤0.05 and M≥0.7 or ≤−0.7). The M value represents the $\log_2$ ratio of transcript levels. Cell cycle related candidates are indicated in bold and experimentally confirmed targets are underlined.
(PDF)

S7 Table. Genes and 5'-/3'-UTRs displaying decreased expression in 2011*ecpR1* versus Rm2011 wild type growing in MOPS medium (P-value≤0.05 and M≥0.7 or ≤−0.7). The M value represents the $\log_2$ ratio of transcript levels. Cell cycle related candidates are denoted in bold and experimentally confirmed targets are underlined.
(PDF)

S8 Table. Genes and 5'-/3'-UTRs displaying increased expression in 2011*ecpR1* versus Rm2011 wild type growing in MOPS medium (P-value$\leq$0.05 and M$\geq$0.7 or $\leq$–0.7). The M value represents the $\log_2$ ratio of transcript levels. Cell cycle related candidates are indicated in bold.
(PDF)

S9 Table. Genes and 5'-/3'-UTRs displaying decreased expression in 2011*ecpR1* versus Rm2011 wild type growing in MOPSlim medium (P-value$\leq$0.05 and M$\geq$0.7 or $\leq$–0.7). The M value represents the $\log_2$ ratio of transcript levels. Cell cycle related candidates are denoted in bold and experimentally confirmed targets are underlined.
(PDF)

S10 Table. Genes and 5'-/3'-UTRs displaying increased expression in 2011*ecpR1* versus Rm2011 wild type growing in MOPSlim medium (P-value$\leq$0.05 and M$\geq$0.7 or $\leq$–0.7). The M value represents the $\log_2$ ratio of transcript levels. Cell cycle related candidates are indicated in bold and experimentally confirmed targets are underlined.
(PDF)

S11 Table. Bacterial strains and plasmids used in this study.
(PDF)

S12 Table. Oligonucleotides used in this study.
(PDF)

S1 Fig. Secondary structure and predicted interaction domains of EcpR1. (A) Secondary structure of the EcpR1 171 nt full length variant with a minimum free energy of -82.10 kcal/mol. Nucleotide positions relative to the first 5'-end are shown. SL, stem loop domain. The 13 nt region predicted to bind the *gcrA* mRNA is boxed. Estimated 5'- and 3'-ends of the four different EcpR1 variants are mapped below. **(B)** Visualization of the predicted interaction domains in the EcpR1 full length sequence. The density plot shows the relative frequency of a specific EcpR1 nucleotide position participating in the top predicted target interactions (P$\leq$0.002). The alignments are shown for the top 20 targets in the EcpR1 prediction (*S. meliloti* and seven closely related *Rhizobiaceae*). The schematic alignment of homologous sRNAs and targets shows the predicted interaction domains: aligned regions are displayed in grey, gaps in white, and predicted interaction regions in different colors. The *S. meliloti* locus tag and gene name (N/A, not available) of the predicted targets are given on the right. **(C)** Predicted EcpR1 binding sites BS1 to BS5 of the *dnaA* mRNA. Nucleotide exchanges in the predicted binding sites BS3 to BS5 that were carried out in different *dnaA* reporter constructs are indicated by arrows and confirmed interactions are shown in bold (see results in Fig 5F). The predicted energy score (E) is indicated in kcal/mol.
(TIF)

S2 Fig. Regulation of the *ecpR1* promoter activity. Northern blot detection of the EcpR1 transcript in Rm2011 wild type at different cell densities (OD_{600}) in TY medium and minimal medium (MM) or in Rm4011 strain carrying pSKEcpR1$^+$ 4 hours after induction with IPTG (EcpR1$^+$) **(A)**, in TY medium and 28 days-old mature symbiotic nodules **(B)** and in the 2011Pσ^{70}*ecpR1* strain carrying a mutation in the -10 region of the σ^{70}-type promoter in stationary growing and oxygen depleted bacteria in TY medium **(F)**. Plots underneath the Northern blot in (A) represent hybridization signal intensities relative to the level of the EcpR1 101 nt variant in Rm2011 growing in TY rich medium at OD_{600} of 0.6, which has been normalized to 1. Promoter alignment of the EcpR1–100 region in different *Sinorhizobium* strains **(C)**. RNAseq-detected EcpR1 5'-ends in Rm2011 are depicted by arrows and the

predicted σ^{70}- and σ^{54}-dependent promoters are underlined. Nucleotide positions are numbered relative to the 5′1 end. Highly and weakly conserved nucleotides are represented as red or blue letters, respectively. Promoter consensus sequences derived from *S. meliloti* 1021 and fragments included in the *ecpR1* transcriptional fusions are indicated below. Means of relative fluorescence intensity values at different cell densities of Rm2011/pP$_{ecpR1}$_5'2 grown in TY **(D)** and of Rm2011 harbouring the corresponding pP$_{ecpR1}$ and the empty vector control, *pleD* or *divK* overexpression plasmids grown in TY medium supplemented with IPTG **(G)**. The standard deviation represents at least three independent measurements of three double transconjugants grown in six independent cultures. Specific activities were normalized to the levels of the stationary phase cultures (OD_{600} of 2.8) (D) or to the cultures lacking IPTG (G) to yield percent relative fluorescence (%F). **(E)** Fluorescence microscopy of exponential and stationary phase Rm2011 cells carrying pP*ecpR1*_5'-2 in TY medium. Bars denote 2 μm.
(TIF)

S3 Fig. Effect of *ecpR1* overexpression on *S. meliloti* generation time and recovery from stationary phase. (A) Time lapse microscopy of Rm4011 cells overexpressing either the control RNA gene SmelC812 (Control$^+$) or *ecpR1* (EcpR1$^+$) after addition of IPTG. Bars denote 2 μm. Cell numbers are indicated in brackets. Doubling times of ~2 hours and ~4 hours were determined for the control RNA gene and *ecpR1* overexpressing strains, respectively. **(B)** Abundance of normal-sized and elongated cells in EcpR1$^+$ cultures treated with IPTG for 30 hours and proportion of stationary cells that resumed growth after washing of cells and transfer to fresh medium lacking the inductor. Proportions were determined by time-lapse microscopy (n = 500). **(C)** Colony forming units (CFUs) of indicated strains after 3 cycles of re-growing on TY medium supplemented with IPTG for 48 h.
(TIF)

S4 Fig. *ecpR1* overexpression-induced phenotype in various α-proteobacteria harbouring an *ecpR1* homolog. Cell morphology **(A)** and DNA content **(B)** of different species overproducing either the control RNA SmelC812 (Control$^+$) or EcpR1 (EcpR1$^+$) 20 hours after addition of IPTG. Bars denote 2 μm.
(TIF)

S5 Fig. 2011*ecpR1* growth and symbiotic phenotypes. Growth rates of Rm2011 and 2011*ecpR1* were compared in TY rich medium at 30°C **(A)**, 45°C **(B)** or after adding 0.4 mM of NaCl **(C)**. Cell viability (CFU/ml) of these two strains was compared after adding 10 mM H_2O_2 to logarithmic cultures in TY for 30 minutes or after growing in defined carbon-limited minimal medium (mannitol 2 g l^{-1}) for 72 h **(D)**. Error bars indicate the standard deviation of at least two replicates. Symbiotic phenotype of *M. truncatula* inoculated with 2011 wild type or 2011*ecpR1*. Time course of *S. meliloti*-induced nodule production **(E)**. Percentage of plants developing nodules **(F)**, and showing a Fix$^+$ phenotype **(G)**. Shoot length of plants growing in the absence of nitrogen 30 days after inoculation with *S. meliloti* 2011, 2011*ecpR1*, and uninoculated (control) **(H)**. Error bars indicate the standard error. All samples were collected from the same experiment (20 plants). Nodulation assays were repeated three times with similar results.
(TIF)

S6 Fig. Fitness of Rm2011 wild type vs. 2011*ecpR1* mutant. Representative fluorescence microscopy images **(A-B)** and means of eGFP:mCherry fluorescence ratios **(C-D)** of 2011 *mCherry* mixed with either 2011*egfp* or 2011*ecpR1 egfp* cell cultures at a 1:1 ratio in MOPS **(A, C)** or MOPSlim media **(B, D)** at the indicated time points. Every 7 days the mixed

population was diluted 1000-fold in fresh media. Standard deviation represents three determinations of three independent cultures. Bars denote 2 μm.
(TIF)

S7 Fig. Target candidates *ctrA*, *minD*, *pleC* and *ftsZ* are not regulated by EcpR1 in *S. meliloti* 4011. Predicted thermodynamically favored antisense interaction regions in *ctrA* **(A)**, *minD* **(B)**, *pleC* **(C)** and *ftsZ1* **(D)** mRNAs, schematic representations of translation fusions to *egfp*, and fluorescence measurements mediated by these constructs in Rm4011*ecpR1* carrying pSKEcpR1$^+$ or pSKControl$^+$. Numbers denote positions relative to the AUG start codon of the mRNA and the second 5'-end of EcpR1. The predicted energy score (E) is indicated in kcal/mol. The standard deviation represents at least three independent determinations of three double transconjugants grown in six independent cultures.
(TIF)

S8 Fig. Target candidates *divJ*, and *divK* are not regulated by EcpR1 in *S. meliloti* 4011. Predicted thermodynamically favored antisense interaction regions in *divJ* **(A)**, SMc00888 **(B)**, and *divK* **(C)** mRNAs, schematic representations of translation fusions to *egfp*, and fluorescence measurements mediated by these constructs in Rm4011*ecpR1* carrying pSKEcpR1$^+$ or pSKControl$^+$. Numbers denote positions relative to the AUG start codon of the mRNA and the second 5'-end of EcpR1. The predicted energy score (E) is indicated in kcal/mol. The standard deviation represents at least three independent determinations of three double transconjugants grown in six independent cultures.
(TIF)

S9 Fig. Molecular and functional characterization of EcpR1 variants starting from the second 5'-end (EcpR1$_{5'2}$), or carrying 1 and 3 nucleotide exchanges in the predicted interaction region (EcpR1-1 and EcpR1-3). Northern blot detection **(A)**, DNA content **(B)** and morphological phenotype **(C)** of Rm4011*ecpR1* overexpressing *ecpR1*$_{5'2}$, *ecpR1-1* and *ecpR1-3*, or control SmelC812. The bar represents 2 μm. **(D, F)** Predicted duplexes between EcpR1 and the *gcrA* mRNA. Nucleotide exchanges in EcpR1 and the *gcrA* mutant variants are denoted in bold. **(E, G)** Fluorescence measurements of 4011*ecpR1* co-transformed with *ecpR1*, *ecpR1-1*, *ecpR1-3*, or control SmelC812 overexpression plasmids and the indicated reporter plasmids. The standard deviation represents at least three independent determinations of three double transconjugants grown in six independent cultures.
(TIF)

Acknowledgments

We thank Barbara Herte and Andreas Kautz for technical assistance in the transcriptome studies, Matthew McIntosh for providing plasmid pPHUtrap and design of the *sinR*-P$_{sinI}$ indirect overexpression strategy, Jan Philip Schlüter and Lars-Ole Loehr for providing plasmids pSGControl$^+$ and pSGEcpR1$^+$, and Oliver Schauer and Pornsri Charoenpanich for providing plasmids pKOSm and pKOSe, respectively, along with other members of the Becker laboratory, for helpful discussions.

Author Contributions

Conceived and designed the experiments: AB MR. Performed the experiments: MR BF PRW. Analyzed the data: MR BF AB. Wrote the paper: AB MR.

References

1. Deng G, Sui G (2013) Noncoding RNA in oncogenesis: a new era of identifying key players. Int J Mol Sci 14: 18319–18349. doi: 10.3390/ijms140918319 PMID: 24013378
2. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136: 215–233. doi: 10.1016/j.cell.2009.01.002 PMID: 19167326
3. He L, He X, Lim LP, de Stanchina E, Xuan Z, et al. (2007) A microRNA component of the p53 tumour suppressor network. Nature 447: 1130–1134. PMID: 17554337
4. Combier JP, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, et al. (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. Genes Dev 20: 3084–3088. PMID: 17114582
5. Bustos-Sanmamed P, Bazin J, Hartmann C, Crespi M, Lelandais-Brière C (2013) Small RNA pathways and diversity in model legumes: lessons from genomics. Front Plant Sci 4: 236. doi: 10.3389/fpls.2013.00236 PMID: 23847640
6. Ding DQ, Okamasa K, Yamane M, Tsutsumi C, Haraguchi T, et al. (2012) Meiosis-specific noncoding RNA mediates robust pairing of homologous chromosomes in meiosis. Science 336: 732–736. doi: 10.1126/science.1219518 PMID: 22582262
7. Yamashita A, Watanabe Y, Nukina N, Yamamoto M (1998) RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. Cell 95: 115–123. PMID: 9778252
8. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. Cell 136: 615–628. doi: 10.1016/j.cell.2009.01.043 PMID: 19239884
9. Storz G, Vogel J, Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43: 880–891. doi: 10.1016/j.molcel.2011.08.022 PMID: 21925377
10. Jonas K (2014) To divide or not to divide: control of the bacterial cell cycle by environmental cues. Curr Opin Microbiol 18: 54–60. doi: 10.1016/j.mib.2014.02.006 PMID: 24631929
11. Tsokos CG, Laub MT (2012) Polarity and cell fate asymmetry in *Caulobacter crescentus*. Curr Opin Microbiol 15: 744–750. doi: 10.1016/j.mib.2012.10.011 PMID: 23146566
12. Curtis PD, Brun YV (2010) Getting in the loop: regulation of development in *Caulobacter crescentus*. Microbiol Mol Biol Rev 74: 13–41. doi: 10.1128/MMBR.00040-09 PMID: 20197497
13. Collier J, Murray SR, Shapiro L (2006) DnaA couples DNA replication and the expression of two cell cycle master regulators. EMBO J 25: 346–356. PMID: 16395331
14. McAdams HH, Shapiro L (2009) System-level design of bacterial cell cycle control. FEBS Lett 583: 3984–3991. doi: 10.1016/j.febslet.2009.09.030 PMID: 19766635
15. Laub MT, Chen SL, Shapiro L, McAdams HH (2002) Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. Proc Natl Acad Sci U S A 99: 4632–4637. PMID: 11930012
16. Marczynski GT, Shapiro L (2002) Control of chromosome replication in *Caulobacter crescentus*. Annu Rev Microbiol 56: 625–656. PMID: 12142494
17. Biondi EG, Reisinger SJ, Skerker JM, Arif M, Perchuk BS, et al. (2006) Regulation of the bacterial cell cycle by an integrated genetic circuit. Nature 444: 899–904. PMID: 17136100
18. Berghoff BA, Glaeser J, Sharma CM, Vogel J, Klug G (2009) Photooxidative stress-induced and abundant small RNAs in Rhodobacter sphaeroides. Mol Microbiol 74: 1497–1512. doi: 10.1111/j.1365-2958.2009.06949.x PMID: 19906181
19. Landt SG, Abeliuk E, McGrath PT, Lesley JA, McAdams HH, et al. (2008) Small non-coding RNAs in *Caulobacter crescentus*. Mol Microbiol 68: 600–614. doi: 10.1111/j.1365-2958.2008.06172.x PMID: 18373523
20. Becker A, Overlöper A, Schlüter JP, Reinkensmeier J, Robledo M, et al. (2014) Riboregulation in plant-associated α-proteobacteria. RNA Biol 11.
21. Schlüter JP, Reinkensmeier J, Daschkey S, Evguenieva-Hackenberg E, Janssen S, et al. (2010) A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. BMC Genomics 11: 245. doi: 10.1186/1471-2164-11-245 PMID: 20398411
22. Schlüter JP, Reinkensmeier J, Barnett MJ, Lang C, Krol E, et al. (2013) Global mapping of transcription start sites and promoter motifs in the symbiotic α-proteobacterium *Sinorhizobium meliloti* 1021. BMC Genomics 14: 156. doi: 10.1186/1471-2164-14-156 PMID: 23497287
23. Gibson KE, Kobayashi H, Walker GC (2008) Molecular determinants of a symbiotic chronic infection. Annu Rev Genet 42: 413–441. doi: 10.1146/annurev.genet.42.110807.091427 PMID: 18983260
24. Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC (2007) How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. Nat Rev Microbiol 5: 619–633. PMID: 17632573

25. De Nisco NJ, Abo RP, Wu CM, Penterman J, Walker GC (2014) Global analysis of cell cycle gene expression of the legume symbiont *Sinorhizobium meliloti*. Proc Natl Acad Sci U S A.

26. Brilli M, Fondi M, Fani R, Mengoni A, Ferri L, et al. (2010) The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. BMC Syst Biol 4: 52. doi: 10.1186/1752-0509-4-52 PMID: 20426835

27. Reinkensmeier R, Schluter JP, Giegerich R, Becker A (2011) Conservation and occurrence of trans-encoded sRNAs in the Rhizobiales. Genes 2: 925–956 doi: 10.3390/genes2040925 PMID: 24710299

28. Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, et al. (2013) Comparative genomics boosts target prediction for bacterial small RNAs. Proc Natl Acad Sci U S A 110: E3487–3496. doi: 10.1073/pnas.1303248110 PMID: 23980183

29. Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J (2008) Small RNA binding to 5' mRNA coding region inhibits translational initiation. Mol Cell 32: 827–837. doi: 10.1016/j.molcel.2008.10.027 PMID: 19111662

30. Sharma CM, Darfeuille F, Plantinga TH, Vogel J (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Genes Dev 21: 2804–2817. PMID: 17974919

31. del Val C, Rivas E, Torres-Quesada O, Toro N, Jiménez-Zurdo JI (2007) Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. Mol Microbiol 66: 1080–1091. PMID: 17971083

32. Pini F, Frage B, Ferri L, De Nisco NJ, Mohapatra SS, et al. (2013) The DivJ, CbrA and PleC system controls DivK phosphorylation and symbiosis in *Sinorhizobium meliloti*. Mol Microbiol 90: 54–71. doi: 10.1111/mmi.12347 PMID: 23909720

33. Ulvé VM, Sevin EW, Chéron A, Barloy-Hubler F (2007) Identification of chromosomal alpha-proteobacterial small RNAs by comparative genome analysis and detection in *Sinorhizobium meliloti* strain 1021. BMC Genomics 8: 467. PMID: 18093320

34. Cheng J, Sibley CD, Zaheer R, Finan TM (2007) A *Sinorhizobium meliloti minE* mutant has an altered morphology and exhibits defects in legume symbiosis. Microbiology 153: 375–387. PMID: 17259609

35. Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, et al. (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic Acids Res 42: W119–123. doi: 10.1093/nar/gku359 PMID: 24838564

36. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008) The Vienna RNA websuite. Nucleic Acids Res 36: W70–74. doi: 10.1093/nar/gkn188 PMID: 18424795

37. Valverde C, Livny J, Schlüter JP, Reinkensmeier J, Becker A, et al. (2008) Prediction of *Sinorhizobium meliloti* sRNA genes and experimental detection in strain 2011. BMC Genomics 9: 416. doi: 10.1186/1471-2164-9-416 PMID: 18793445

38. Krol E, Becker A (2011) ppGpp in *Sinorhizobium meliloti*: biosynthesis in response to sudden nutritional downshifts and modulation of the transcriptome. Mol Microbiol 81: 1233–1254. doi: 10.1111/j.1365-2958.2011.07752.x PMID: 21696469

39. Sibley CD, MacLellan SR, Finan T (2006) The *Sinorhizobium meliloti* chromosomal origin of replication. Microbiology 152: 443–455. PMID: 16436432

40. Latch JN, Margolin W (1997) Generation of buds, swellings, and branches instead of filaments after blocking the cell cycle of *Rhizobium meliloti*. J Bacteriol 179: 2373–2381. PMID: 9079925

41. Kahng LS, Shapiro L (2001) The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* is essential, and its activity is cell cycle regulated. J Bacteriol 183: 3065–3075. PMID: 11325934

42. Bahlawane C, McIntosh M, Krol E, Becker A (2008) *Sinorhizobium meliloti* regulator MucR couples exopolysaccharide synthesis and motility. Mol Plant Microbe Interact 21: 1498–1509. doi: 10.1094/MPMI-21-11-1498 PMID: 18842098

43. Fields AT, Navarrete CS, Zare AZ, Huang Z, Mostafavi M, et al. (2012) The conserved polarity factor *podJ1* impacts multiple cell envelope-associated functions in *Sinorhizobium meliloti*. Mol Microbiol 84: 892–920. doi: 10.1111/j.1365-2958.2012.08064.x PMID: 22553970

44. Torres-Quesada O, Millán V, Nisa-Martínez R, Bardou F, Crespi M, et al. (2013) Independent activity of the homologous small regulatory RNAs AbcR1 and AbcR2 in the legume symbiont *Sinorhizobium meliloti*. PLoS One 8: e68147. doi: 10.1371/journal.pone.0068147 PMID: 23869210

45. Giacomini A, Ollero FJ, Squartini A, Nuti MP (1994) Construction of multipurpose gene cartridges based on a novel synthetic promoter for high-level gene expression in gram-negative bacteria. Gene 144: 17–24. PMID: 8026755

46. Greif D, Pobigaylo N, Frage B, Becker A, Regtmeier J, et al. (2010) Space- and time-resolved protein dynamics in single bacterial cells observed on a chip. J Biotechnol 149: 280–288. doi: 10.1016/j.jbiotec.2010.06.003 PMID: 20599571

47. Galibert F, Finan TM, Long SR, Puhler A, Abola P, et al. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293: 668–672. PMID: 11474104

48. Gottesman S, Storz G (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb Perspect Biol 3.

49. Rice JB, Vanderpool CK (2011) The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes. Nucleic Acids Res 39: 3806–3819. doi: 10.1093/nar/gkq1219 PMID: 21245045

50. Desnoyers G, Morissette A, Prévost K, Massé E (2009) Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. EMBO J 28: 1551–1561. doi: 10.1038/emboj.2009.116 PMID: 19407815

51. Majdalani N, Cunning C, Sledjeski D, Elliott T, Gottesman S (1998) DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. Proc Natl Acad Sci U S A 95: 12462–12467. PMID: 9770508

52. Guillier M, Gottesman S (2008) The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. Nucleic Acids Res 36: 6781–6794. doi: 10.1093/nar/gkn742 PMID: 18953042

53. Brennan RG, Link TM (2007) Hfq structure, function and ligand binding. Curr Opin Microbiol 10: 125–133. PMID: 17395525

54. Folichon M, Arluison V, Pellegrini O, Huntzinger E, Régnier P, et al. (2003) The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. Nucleic Acids Res 31: 7302–7310. PMID: 14654705

55. Torres-Quesada O, Reinkensmeier J, Schlüter JP, Robledo M, Peregrina A, et al. (2014) Genome-wide profiling of Hfq-binding RNAs uncovers extensive post-transcriptional rewiring of major stress response and symbiotic regulons in *Sinorhizobium meliloti*. RNA Biol 11.

56. Torres-Quesada O, Oruezabal RI, Peregrina A, Jofré E, Lloret J, et al. (2010) The *Sinorhizobium meliloti* RNA chaperone Hfq influences central carbon metabolism and the symbiotic interaction with alfalfa. BMC Microbiol 10: 71. doi: 10.1186/1471-2180-10-71 PMID: 20205931

57. Gao M, Barnett MJ, Long SR, Teplitski M (2010) Role of the *Sinorhizobium meliloti* global regulator Hfq in gene regulation and symbiosis. Mol Plant Microbe Interact 23: 355–365. doi: 10.1094/MPMI-23-4-0355 PMID: 20192823

58. Morita T, Maki K, Aiba H (2005) RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. Genes Dev 19: 2176–2186. PMID: 16166379

59. Baumgardt K, Charoenpanich P, McIntosh M, Schikora A, Stein E, et al. (2014) RNase E affects the expression of the acyl-homoserine lactone synthase gene *sinI* in *Sinorhizobium meliloti*. J Bacteriol 196: 1435–1447. doi: 10.1128/JB.01471-13 PMID: 24488310

60. Fioravanti A, Fumeaux C, Mohapatra SS, Bompard C, Brilli M, et al. (2013) DNA binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in *Caulobacter crescentus* and other *Alphaproteobacteria*. PLoS Genet 9: e1003541. doi: 10.1371/journal.pgen.1003541 PMID: 23737758

61. Jonas K, Liu J, Chien P, Laub MT (2013) Proteotoxic stress induces a cell-cycle arrest by stimulating Lon to degrade the replication initiator DnaA. Cell 154: 623–636. doi: 10.1016/j.cell.2013.06.034 PMID: 23911325

62. Tétart F, Bouché JP (1992) Regulation of the expression of the cell-cycle gene *ftsZ* by DicF antisense RNA. Division does not require a fixed number of FtsZ molecules. Mol Microbiol 6: 615–620. PMID: 1372677

63. Dadzie I, Xu S, Ni B, Zhang X, Zhang H, et al. (2013) Identification and characterization of a *cis*-encoded antisense RNA associated with the replication process of *Salmonella enterica* serovar Typhi. PLoS One 8: e61308. doi: 10.1371/journal.pone.0061308 PMID: 23637809

64. Yu YT, Yuan X, Velicer GJ (2010) Adaptive evolution of an sRNA that controls *Myxococcus* development. Science 328: 993. doi: 10.1126/science.1187200 PMID: 20489016

65. Grieshaber NA, Grieshaber SS, Fischer ER, Hackstadt T (2006) A small RNA inhibits translation of the histone-like protein Hc1 in Chlamydia trachomatis. Mol Microbiol 59: 541–550. PMID: 16390448

66. Tattersall J, Rao GV, Runac J, Hackstadt T, Grieshaber SS, et al. (2012) Translation inhibition of the developmental cycle protein HctA by the small RNA IhtA is conserved across *Chlamydia*. PLoS One 7: e47439. doi: 10.1371/journal.pone.0047439 PMID: 23071807

67. Chant EL, Summers DK (2007) Indole signalling contributes to the stable maintenance of *Escherichia coli* multicopy plasmids. Mol Microbiol 63: 35–43. PMID: 17163976

68. Papenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JC, et al. (2006) SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. Mol Microbiol 62: 1674–1688. PMID: 17427289

69. Penterman J, Abo RP, De Nisco NJ, Arnold MF, Longhi R, et al. (2014) Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. Proc Natl Acad Sci U S A.

70. Roux B, Rodde N, Jardinaud MF, Timmers T, Sauviac L, et al. (2014) An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J 77: 817–837. doi: 10.1111/tpj.12442 PMID: 24483147

71. Braun RE, O'Day K, Wright A (1985) Autoregulation of the DNA replication gene *dnaA* in *E*. *coli* K-12. Cell 40: 159–169. PMID: 2981626

72. Holtzendorff J, Hung D, Brende P, Reisenauer A, Viollier PH, et al. (2004) Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. Science 304: 983–987. PMID: 15087506

73. Zweiger G, Shapiro L (1994) Expression of *Caulobacter dnaA* as a function of the cell cycle. J Bacteriol 176: 401–408. PMID: 8288535

74. Desnoyers G, Bouchard MP, Massé E (2013) New insights into small RNA-dependent translational regulation in prokaryotes. Trends Genet 29: 92–98. doi: 10.1016/j.tig.2012.10.004 PMID: 23141721

75. Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC, et al. (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. Mol Microbiol 84: 428–445. doi: 10.1111/j.1365-2958.2012.08031.x PMID: 22458297

76. Rice JB, Balasubramanian D, Vanderpool CK (2012) Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA. Proc Natl Acad Sci U S A 109: E2691–2698. doi: 10.1073/pnas.1207927109 PMID: 22988087

77. Beringer JE (1974) R factor transfer in *Rhizobium leguminosarum*. J Gen Microbiol 84: 188–198. PMID: 4612098

78. Zhan HJ, Lee CC, Leigh JA (1991) Induction of the second exopolysaccharide (EPSb) in *Rhizobium meliloti* SU47 by low phosphate concentrations. J Bacteriol 173: 7391–7394. PMID: 1938929

79. Robledo M, Jiménez-Zurdo JI, Velázquez E, Trujillo ME, Zurdo-Piñeiro JL, et al. (2008) Rhizobium cellulase CelC2 is essential for primary symbiotic infection of legume host roots. Proc Natl Acad Sci U S A 105: 7064–7069. doi: 10.1073/pnas.0802547105 PMID: 18458328

80. Schäfer A, Tauch A, Jäger W, Kalinowski J, Thierbach G, et al. (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene 145: 69–73. PMID: 8045426

81. McIntosh M, Meyer S, Becker A (2009) Novel *Sinorhizobium meliloti* quorum sensing positive and negative regulatory feedback mechanisms respond to phosphate availability. Mol Microbiol 74: 1238–1256. doi: 10.1111/j.1365-2958.2009.06930.x PMID: 19889097

82. Charoenpanich P, Meyer S, Becker A, McIntosh M (2013) Temporal expression program of quorum sensing-based transcription regulation in *Sinorhizobium meliloti*. J Bacteriol 195: 3224–3236. doi: 10.1128/JB.00234-13 PMID: 23687265

83. Hübner P, Willison JC, Vignais PM, Bickle TA (1991) Expression of regulatory *nif* genes in *Rhodobacter capsulatus*. J Bacteriol 173: 2993–2999. PMID: 1902215

84. Serrania J, Vorhölter FJ, Niehaus K, Pühler A, Becker A (2008) Identification of *Xanthomonas campestris pv*. *campestris* galactose utilization genes from transcriptome data. J Biotechnol 135: 309–317. doi: 10.1016/j.jbiotec.2008.04.011 PMID: 18538881

85. Dondrup M, Albaum SP, Griebel T, Henckel K, Jünemann S, et al. (2009) EMMA 2—a MAGE-compliant system for the collaborative analysis and integration of microarray data. BMC Bioinformatics 10: 50. doi: 10.1186/1471-2105-10-50 PMID: 19200358

86. Becker A, Bergès H, Krol E, Bruand C, Rüberg S, et al. (2004) Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. Mol Plant Microbe Interact 17: 292–303. PMID: 15000396

87. Darty K, Denise A, Ponty Y (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics 25: 1974–1975. doi: 10.1093/bioinformatics/btp250 PMID: 19398448

88. Smith C, Heyne S, Richter AS, Will S, Backofen R (2010) Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. Nucleic Acids Res 38: W373–377. doi: 10.1093/nar/gkq316 PMID: 20444875

89. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. doi: 10.1016/j.jneumeth.2014.06.019 PMID: 24970579

8 Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*

Erik Holmqvist, **Patrick R. Wright**, Lei Li, Thorsten Bischler, Lars Barquist, Richard Reinhardt, Rolf Backofen and Jörg Vogel (2016) **The EMBO Journal**, 35, 991-1011.
DOI: 10.15252/embj.201593360

Personal contribution

I designed and implemented the peak calling approach for this project. Furthermore, I performed the analysis showing how Hfq CLIP data can improve sRNA target prediction.

Patrick R. Wright

The following co-authors confirm the above stated contribution.

Prof. Dr. Jörg Vogel

Dr. Erik Holmqvist

Published online: April 4, 2016

Resource

SOURCE DATA TRANSPARENT PROCESS OPEN ACCESS

THE EMBO JOURNAL

Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*

Erik Holmqvist[1], Patrick R Wright[2], Lei Li[1], Thorsten Bischler[1], Lars Barquist[1], Richard Reinhardt[3], Rolf Backofen[2,4,*] & Jörg Vogel[1,**]

Abstract

The molecular roles of many RNA-binding proteins in bacterial post-transcriptional gene regulation are not well understood. Approaches combining *in vivo* UV crosslinking with RNA deep sequencing (CLIP-seq) have begun to revolutionize the transcriptome-wide mapping of eukaryotic RNA-binding protein target sites. We have applied CLIP-seq to chart the target landscape of two major bacterial post-transcriptional regulators, Hfq and CsrA, in the model pathogen *Salmonella* Typhimurium. By detecting binding sites at single-nucleotide resolution, we identify RNA preferences and structural constraints of Hfq and CsrA during their interactions with hundreds of cellular transcripts. This reveals 3′-located Rho-independent terminators as a universal motif involved in Hfq–RNA interactions. Additionally, Hfq preferentially binds 5′ to sRNA-target sites in mRNAs, and 3′ to seed sequences in sRNAs, reflecting a simple logic in how Hfq facilitates sRNA–mRNA interactions. Importantly, global knowledge of Hfq sites significantly improves sRNA-target predictions. CsrA binds AUGGA sequences in apical loops and targets many *Salmonella* virulence mRNAs. Overall, our generic CLIP-seq approach will bring new insights into post-transcriptional gene regulation by RNA-binding proteins in diverse bacterial species.

Keywords CLIP; CsrA; Hfq; non-coding RNA; peak calling; post-transcriptional control; small RNA; terminator; translation

Subject Categories Methods & Resources; Microbiology, Virology & Host Pathogen Interaction; RNA Biology

DOI 10.15252/embj.201593360 | Received 9 November 2015 | Revised 25 February 2016 | Accepted 26 February 2016

Introduction

The fate of RNA molecules in the cell is largely determined at the post-transcriptional level by RNA–protein interactions. RNA-binding proteins (RBPs) are responsible for essential traits such as RNA stability, structure, translatability, export, and localization. Recent screens in human cells have suggested that the number of proteins with RNA-binding properties may be vastly underestimated (Baltz *et al*, 2012; Castello *et al*, 2012; Kramer *et al*, 2014), prompting new systematic searches for RBPs in many eukaryotic systems (Ascano *et al*, 2013). By comparison, our knowledge of the scope and binding preferences of prokaryotic RBPs is lagging behind eukaryotic systems, and new approaches are needed to fully elucidate the roles of RBPs in post-transcriptional control in bacterial pathogens (Barquist & Vogel, 2015). That is, although the structural details of the interactions of many positively and negatively acting proteins with DNA have been established, the paucity of understanding regarding RBPs has been holding back the field of bacterial gene regulation.

Salmonella enterica serovar Typhimurium is a widely studied food-borne bacterial pathogen that invades and replicates in many different eukaryotic host cells. Over the past decade, *Salmonella* has become a bacterial model organism to study post-transcriptional regulation by small regulatory RNAs (sRNAs) and two associated RBPs, Hfq and CsrA (Vogel, 2009; Hébrard *et al*, 2012; Westermann *et al*, 2016). Transcriptomic and RNA co-immunoprecipitation (coIP) analyses have suggested that Hfq and CsrA play global roles in the regulation of *Salmonella* virulence genes (Lawhon *et al*, 2003; Sittka *et al*, 2008; Ansong *et al*, 2009), but precisely how and where these proteins bind cellular transcripts *in vivo* remains to be fully understood.

Hfq is a widely conserved bacterial RBP of the Sm family of proteins which have a ring-like multimeric quaternary structure (Wilusz & Wilusz, 2005). In the Gram-negative bacteria *Salmonella* and *Escherichia coli*, coIP studies have predicted interactions of Hfq

1 Institute for Molecular Infection Biology, University of Würzburg, Würzburg, Germany
2 Bioinformatics Group, Department of Computer Science, Albert Ludwig University Freiburg, Freiburg, Germany
3 Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, Germany
4 BIOSS Centre for Biological Signaling Studies, University of Freiburg, Freiburg, Germany
*Corresponding author. Tel: +49 761 203 7460; E-mail: backofen@informatik.uni-freiburg.de
**Corresponding author. Tel: +49 931 318 2576; E-mail: joerg.vogel@uni-wuerzburg.de

Published online: April 4, 2016

with hundreds of sRNAs and an excess of one thousand mRNAs (Chao *et al*, 2012; Zhang *et al*, 2013; Bilusic *et al*, 2014). By helping sRNAs to regulate target mRNAs, Hfq modulates a variety of physiological traits including phosphosugar detoxification (Rice *et al*, 2012; Papenfort *et al*, 2013), catabolite repression (Beisel *et al*, 2012), envelope stress (Figueroa-Bossi *et al*, 2006; Gogol *et al*, 2011; Guo *et al*, 2014; Chao & Vogel, 2016), metal homeostasis (Desnoyers & Masse, 2012; Coornaert *et al*, 2013), biofilm formation (Holmqvist *et al*, 2010; Jørgensen *et al*, 2012; Mika *et al*, 2012; Thomason *et al*, 2012), motility (De Lay & Gottesman, 2012), and virulence (Sittka *et al*, 2007; Koo *et al*, 2011; Westermann *et al*, 2016). In pathogenic *Vibrio* species, Hfq and sRNAs regulate similarly complex traits, for example, quorum sensing or biofilm formation (Feng *et al*, 2015; Papenfort *et al*, 2015).

Mechanistically, Hfq promotes sRNA–mRNA annealing by increasing the rate of duplex formation (Møller *et al*, 2002; Zhang *et al*, 2002; Lease & Woodson, 2004; Link *et al*, 2009; Fender *et al*, 2010), while at the same time protecting sRNAs from the activity of cellular ribonucleases (Vogel & Luisi, 2011). In addition, Hfq may recruit auxiliary protein factors such as RNase E to promote the decay of target mRNAs (Morita & Aiba, 2011; Bandyra *et al*, 2012).

Structural studies of *Salmonella* Hfq confirmed the homo-hexameric ring model (Sauer & Weichenrieder, 2011). The two faces of the ring, denoted proximal and distal, both bind RNA, but show affinity for different RNA sequences: the proximal face tends to target single-stranded U-rich sequences, whereas the distal face interacts with single-stranded A-rich sequences (Schumacher *et al*, 2002; Mikulecky *et al*, 2004; Link *et al*, 2009). More recently, the rim of the Hfq hexamer has emerged as a third RNA-binding surface which interacts with UA-rich RNA and promotes intermolecular RNA annealing (Updegrove & Wartell, 2011; Sauer *et al*, 2012; Panja *et al*, 2013; Dimastrogiovanni *et al*, 2014). Whereas most of these findings stem from studying Hfq interactions with selected model substrates *in vitro*, details of transcriptome-wide Hfq binding within RNA *in vivo* emerged only recently through a crosslinking-based study in pathogenic *E. coli* (Tree *et al*, 2014). However, while this study captured many known Hfq targets, it generally failed to observe Hfq binding to sRNA 3′ ends, thus contrasting with the emerging mechanistic model from recent biochemical and structural studies whereby Hfq is loaded onto sRNAs via their 3′ located poly(U) stretch (Otaka *et al*, 2011; Sauer & Weichenrieder, 2011; Ishikawa *et al*, 2012; Dimastrogiovanni *et al*, 2014).

CsrA, initially identified as a regulator of carbon storage and glycogen biosynthesis in *E. coli* (Romeo *et al*, 1993), belongs to the large CsrA/Rsm family of RBPs that influence physiology and virulence in numerous pathogenic and non-pathogenic bacteria (Lenz *et al*, 2005; Brencic & Lory, 2009; Heroven *et al*, 2012; Romeo *et al*, 2013; Vakulskas *et al*, 2015). CsrA/Rsm proteins primarily affect translation of mRNAs by binding to 5′ untranslated regions (UTRs). A wealth of genetic, biochemical, and structural data shows that these proteins generally recognize GGA motifs in apical loops of RNA secondary structures (Dubey *et al*, 2005; Duss *et al*, 2014a). Other reported mechanisms of CsrA activity in the cell include promotion of Rho-dependent transcription termination, or mRNA stabilization by masking of RNase E cleavage sites (Yakhnin *et al*, 2013; Figueroa-Bossi *et al*, 2014). CsrA may also govern a large post-transcriptional regulon, as inferred from transcriptomic and RNA co-purification data in *Salmonella* and *E. coli*, respectively (Lawhon *et al*, 2003; Edwards *et al*, 2011).

The CsrA/Rsm proteins are themselves regulated by sRNAs such as CsrB and RsmZ, which contain multiple GGA sites that titrate the protein away from mRNA targets (Liu *et al*, 1997; Weilbacher *et al*, 2003; Valverde *et al*, 2004). Structural studies of one CsrA-like protein revealed a sequential and cooperative assembly of the protein on antagonistic sRNAs (Duss *et al*, 2014b). Antagonists of CsrA activity also include the Hfq-dependent sRNA McaS in *E. coli* (Holmqvist & Vogel, 2013; Jørgensen *et al*, 2013) and a sponge-like mRNA in *Salmonella* (Sterzenbach *et al*, 2013). Again, despite the strong interest in these proteins, the global binding preferences of CsrA/Rsm *in vivo* remain unknown.

Approaches combining *in vivo* crosslinking and RNA deep sequencing have been increasingly used to globally map the cellular RNA ligands and binding sites of eukaryotic RBPs *in vivo* (Darnell, 2010; König *et al*, 2011; Ascano *et al*, 2012). Such methods are now widely used in cell culture, tissues, and even whole animals. The purification of RNA–protein complexes after *in vivo* crosslinking by ultraviolet (UV) light offers several advantages over traditional coIP. Firstly, the UV-induced covalent bonds between protein and RNA survive denaturing conditions, facilitating stringent purification protocols. Secondly, crosslinking enables trimming by ribonucleases to yield protein-protected RNA fragments, pinpointing binding regions with unprecedented resolution. Thirdly, the attachment of a crosslinked peptide to a purified RNA fragment often causes mutations during reverse transcription which identify direct RNA–protein contacts at single-nucleotide resolution (Zhang & Darnell, 2011).

Here, we have employed UV crosslinking of RNA–protein complexes in living bacterial cells, followed by stringent purification and sequencing of crosslinked RNA, to detect transcriptome-wide binding sites of Hfq and CsrA in *Salmonella*. As well as confirming known binding sites at nucleotide resolution, our study identifies a plethora of new sites that reveal the specificities of Hfq and CsrA interactions with their RNA ligands. Our contact maps for Hfq interacting sRNAs and their target mRNAs support a model for Hfq as a mediator of RNA duplex formation and provide new insight into improving sRNA-target prediction. The discovery of CsrA-binding sites in mRNAs shows that CsrA is a direct regulator of *Salmonella* virulence genes.

Results

Selective enrichment of crosslinked RNA ligands

To comprehensively analyze direct targets of RBPs *in vivo*, we established a CLIP-seq protocol for purification of crosslinked RNA–protein complexes from bacterial cells irradiated with UV light (Fig 1A). *Salmonella* strain SL1344 expressing chromosomally FLAG-tagged Hfq was cultured in LB medium to an OD_{600} of 2.0. One half of this culture was then irradiated with UV light while the other half was left untreated. This growth condition activates the invasion genes of *Salmonella*, that is it enabled us to also capture potential Hfq interactions with virulence-associated transcripts. Hfq–RNA complexes were immunoprecipitated in cell lysates with a monoclonal anti-FLAG antibody followed by several stringent washes.

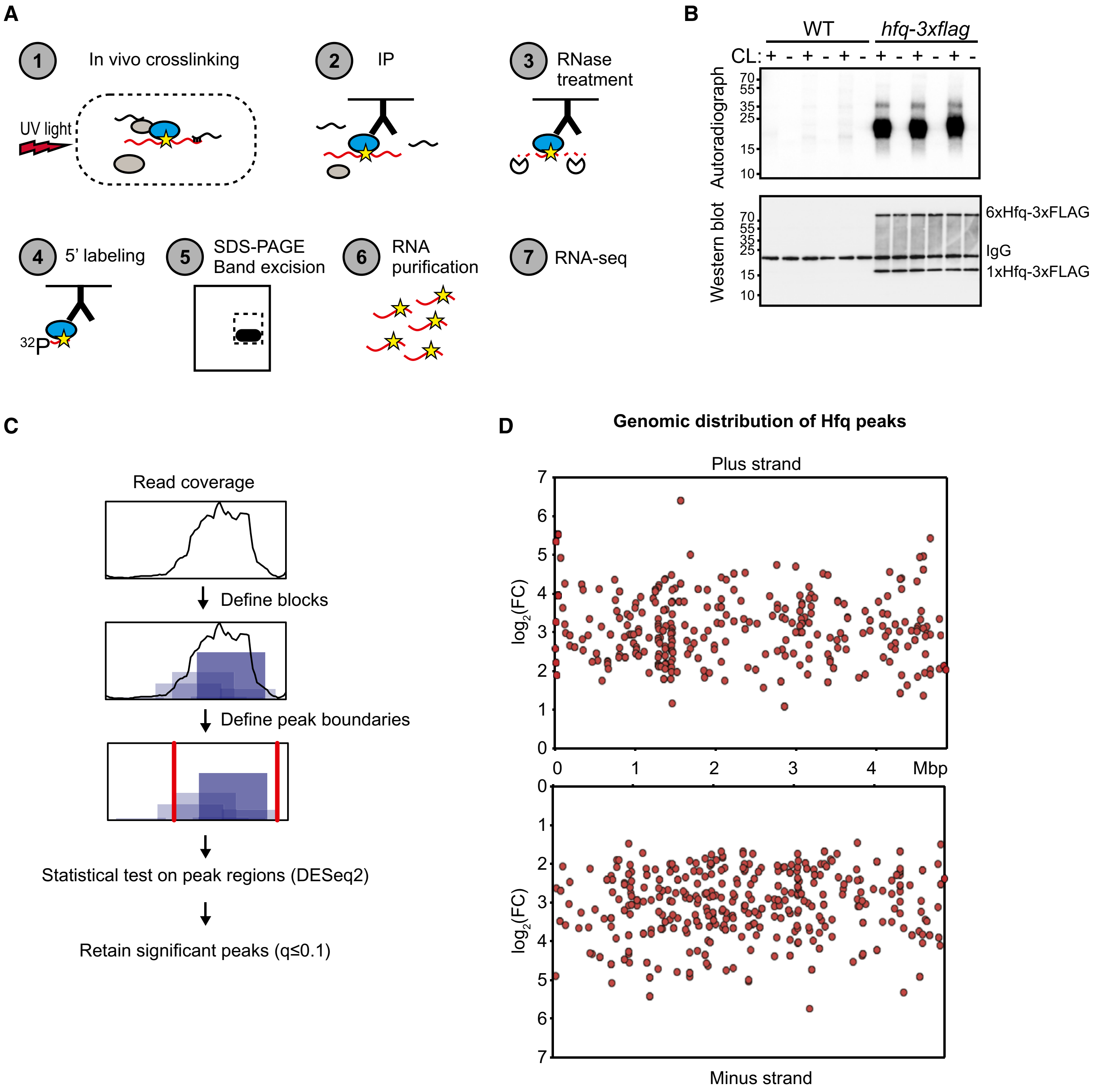


Figure 1. CLIP-seq of Hfq-3xFLAG in *Salmonella*.

A Schematic representation of the CLIP-seq protocol for bacterial RBPs that was established and used in this study. UV: ultraviolet.

B Detection of crosslinked, immunoprecipitated, and radioactively labeled RNA–protein complexes after separation on denaturing SDS–polyacrylamide gels and transfer to nitrocellulose membranes. Radioactive signals were detected by phosphorimaging (top). Detection of Hfq-3xFLAG proteins by Western blot using an anti-FLAG antibody served as a control for successful immunoprecipitation (bottom). CL: crosslinking.

C Schematic representation of binding site determination (peak calling).

D Fold change (*y*-axis) and genomic position (*x*-axis) of Hfq peaks. Mbp: mega basepair.

After on-bead RNase treatment, dephosphorylation, and radioactive labeling of RNA 5′ ends, the complexes were eluted, separated by denaturing SDS–PAGE, and transferred to a membrane. UV irradiation itself did not interfere with protein recovery (as judged by Western blot), but a strong radioactive signal corresponding to bound labeled RNA was detected only in tagged and crosslinked samples, indicating that unspecific RNA–protein interactions were successfully depleted (Fig 1B). RNA–protein complexes from

Published online: April 4, 2016

crosslinked and control samples were extracted from the membrane and treated with proteinase to yield RNA ligands for analysis by Illumina sequencing. The number of sequencing reads obtained for each cDNA library is given in Appendix Fig S1. To avoid biases introduced during library amplification, reads originating from potential PCR duplicates were removed for all downstream analyses.

A very important step in the analysis of CLIP-seq data is peak calling, which is used to differentiate between specific und unspecific binding. Here, two major problems in standard CLIP-seq protocols may confound peak calling approaches. Firstly, in contrast to traditional RNA immunoprecipitation and sequencing (RIP-seq), where comparison to a non-tagged strain or the omission of the antibody serves to control for background noise, CLIP-seq approaches usually lack a standardized negative control. Secondly, in contrast to chromatin immunoprecipitation and DNA sequencing (ChIP-seq), transcript abundance impacts read coverage independent of the affinity of the RBP for a given target. Standard peak callers such as Piranha (Uren *et al*, 2012) assume the majority of sites to be noise, so the sum of all sites can be used to fit a background model. However, this assumption is problematic if the RBP is a ubiquitous binder and the genome size is rather small. Both criteria apply in our case. To overcome these problems, we developed a specific peak calling algorithm able to identify Hfq-binding sites throughout the *Salmonella* transcriptome. The algorithm first divides consecutive reads into blocks and then merges overlapping blocks into peaks (Fig 1C). Subsequently, based on three biological replicates and three control replicates, each peak was tested for significant enrichment in the crosslinked samples versus the non-crosslinked samples using DESeq2 (Love *et al*, 2014). This strategy identified 640 significant ($q \leq 0.1$) Hfq peaks (Table EV1) which are distributed across the *Salmonella* transcriptome (Fig 1D).

As a significant advantage of CLIP-seq over simple coIP, crosslinking-induced mutations narrow RNA–protein contacts down to individual nucleotides (Zhang & Darnell, 2011). Thus, we compared the nature of read mutations that (i) occurred in both mate pairs for each read (to discriminate from sequencing errors), (ii) were exclusively present in libraries from crosslinked cultures, and (iii) overlapped with Hfq peaks (Table EV2). T to C mutations were by far the most common crosslink-specific mutation (Fig 2A), and more than half of the Hfq peaks (347/640) contained at least one crosslink-specific mutation. To provide a better display of peak density, the *Salmonella* chromosome was divided into bins of 2×10^4 basepairs. Plotting peak numbers per bin identified certain chromosomal regions in which the density of Hfq peaks is unusually high (Fig 2B). Interestingly, transcripts from the two major pathogenicity islands, SPI-1 and SPI-2, attract the highest Hfq peak density, supporting the crucial role of Hfq in *Salmonella* virulence (Sittka *et al*, 2007). Dividing the Hfq peaks into different RNA classes shows that the majority map to sRNAs and mRNAs, the two RNA classes previously known to be targets of Hfq (Fig 2C). In summary, combining CLIP-seq with a new peak calling algorithm and identification of crosslinking-induced mutations provides the basis for a detailed investigation of Hfq–RNA interactions.

Hfq binding in mRNAs

To analyze the general distribution of the 551 Hfq-binding sites detected in mRNAs, we performed a meta-gene analysis of Hfq peaks with respect to mRNA start and stop codons (for polycistronic mRNAs, only the start codon of the first cistron and the stop codon of the last cistron was used). The greatest peak densities were found in 5′UTRs and 3′UTRs (Fig 2D) and confirmed—on the level of individual transcripts—previously predicted Hfq activity, for example, in the 5′UTR of *chiP* mRNA which is a target of ChiX sRNA (Figueroa-Bossi *et al*, 2009), or the 3′UTR of *hilD* mRNA encoding a virulence regulator (Lopez-Garrido *et al*, 2014) (Fig 2E and F).

To test whether Hfq recognizes disparate sequences in different parts of mRNAs, we divided the mRNA peaks into those that map to 5′UTRs, CDSs, or 3′UTRs. Using the MEME algorithm (Bailey *et al*, 2015), only the combined 3′UTRs yielded a significant consensus motif (Fig 2G). This motif strongly resembles Rho-independent transcription terminators present at the 3′ end of many bacterial transcripts, namely GC-rich hairpins followed by single-stranded uridine tails (Wilson & von Hippel, 1995). Indeed, we found a strong enrichment of Hfq 3′UTR peaks at predicted Rho-independent terminators that were specific to mRNAs (Fig 2H; all sRNA terminators were excluded from this analysis). Moreover, CMfinder analysis (Yao *et al*, 2006) on the Hfq 3′UTR peaks resulted in a motif comprising a hairpin structure followed by a U-rich sequence, strongly resembling a Rho-independent terminator (Fig EV1), suggesting that Hfq binds to mRNA 3′ ends.

Hfq binding in sRNAs

We next compared our crosslinking data to Hfq-binding sites in well-investigated sRNAs. For example, SgrS was proposed to contain an Hfq-binding module consisting of two distinct binding sites: the poly(U) sequence of the Rho-independent terminator at the very 3′ end of SgrS, and an internal hairpin preceded by a U-rich sequence (Ishikawa *et al*, 2012). In accordance with this, we detected two Hfq peaks within SgrS that mapped to the previously reported binding sites (Fig 3A and B). In addition, the only

Figure 2. Genomic distribution of Hfq-binding sites.

A Percentage of the occurrence of the indicated mutations among all crosslink-specific mutations found within Hfq peaks.

B Hfq peak distribution along the *Salmonella* chromosome divided in bins of 2×10^4 basepairs each. The genomic positions of the pathogenicity islands SPI-1 and SPI-2 are indicated. Mbp: mega basepair.

C Distribution of Hfq peaks among the indicated RNA classes. Numbers in parentheses give the number of called peaks that overlapped with annotations belonging to the respective RNA class.

D Global peak density distribution (meta-gene analysis) around start and stop codons. For this analysis, only those start and stop codons were used that are flanked by a 5′UTR or 3′UTR, respectively. Vertical dashed lines indicate the position of start and stop codons, respectively.

E, F Read coverage at the *chiP* (E) and *hilD* (F) loci in libraries from crosslinked and non-crosslinked samples. Exp: experiment, CL: crosslinking

G Consensus motif generated by MEME using sequences of Hfq peaks mapping to mRNA 3′UTRs.

H Meta-gene analysis of peak distribution around genomic positions of predicted Rho-independent terminators.

▶

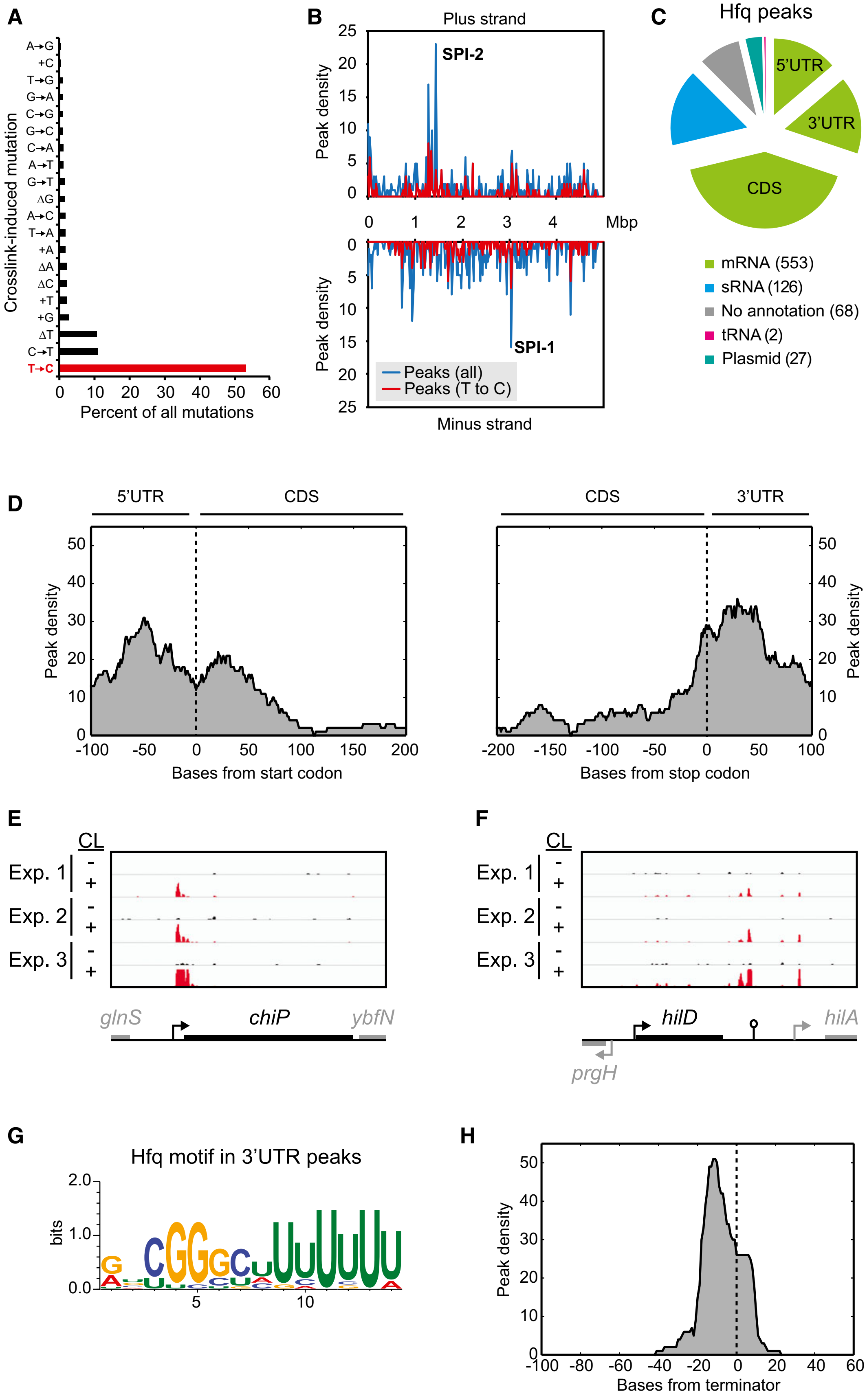

A
A→G
+C
T→G
G→A
C→G
G→C
C→A
A→T
G→T
ΔG
A→C
T→A
+A
ΔA
ΔC
+T
+G
ΔT
C→T
T→C
Crosslink-induced mutation
0 10 20 30 40 50 60
Percent of all mutations
B
Plus strand
SPI-2
Peak density
25 20 15 10 5 0
0 1 2 3 4 Mbp
0 5 10 15 20 25
SPI-1
Peaks (all)
Peaks (T to C)
Minus strand
C
Hfq peaks
5'UTR
3'UTR
CDS
mRNA (553)
sRNA (126)
No annotation (68)
tRNA (2)
Plasmid (27)
D
5'UTR
CDS
CDS
3'UTR
Peak density
50 40 30 20 10 0
-100 -50 0 50 100 150 200
Bases from start codon
-200 -150 -100 -50 0 50 100
Bases from stop codon
E
CL
Exp. 1
Exp. 2
Exp. 3
-
+
glnS
chiP
ybfN
F
CL
Exp. 1
Exp. 2
Exp. 3
-
+
hilD
hilA
prgH
G
Hfq motif in 3'UTR peaks
bits
2.0
1.0
0.0
5
10
H
Peak density
50 40 30 20 10 0
-100 -80 -60 -40 -20 0 20 40 60
Bases from terminator

Figure 2.

Published online: April 4, 2016

crosslink-induced mutations detected in SgrS occur within the above-described U-rich sequences (Fig 3B). Likewise, we compared our crosslinking data with the interactions observed in a co-crystal of *Salmonella* Hfq and the sRNA RydC (Dimastrogiovanni *et al*, 2014). The X-ray crystallization data suggest Hfq interacts with four regions on RydC: the proximal site of Hfq interacts with the U-rich 3′ end of RydC; the rim of Hfq interacts with U23/U24, U46/U47, and the RydC 5′ end (Dimastrogiovanni *et al*, 2014). Out of the eight positions in RydC with crosslinking-induced mutations, seven perfectly fit with the crystal structure (Fig 3D). Mutations were found in the 5′ end of RydC, at positions U23, U24, U46, U47, and in the RydC 3′ end (Fig 3D). Taken together, these examples demonstrate that our crosslinking experiments faithfully capture Hfq–RNA interactions at single-nucleotide resolution, in excellent agreement with published work.

The distribution of Hfq peaks over all sRNA sequences suggests that Hfq may interact with different regions in different sRNAs; however, there is a strong bias for Hfq binding toward sRNA 3′ ends (Fig 3E). As for the 3′UTR-binding motif (Fig 2G), the consensus motif found using MEME in peaks mapping to within sRNAs resembles the 3′ region of a Rho-independent terminator (Fig 3F). Following the demonstration of Hfq interactions with 3′ portions of a few sRNAs (Sauer & Weichenrieder, 2011; Ishikawa *et al*, 2012), our screen provides the first global analysis to suggest that Hfq interacts with the 3′ end of many sRNAs detected under the growth condition studied. Taken together, Rho-independent terminators constitute a general Hfq-binding motif shared by mRNAs and sRNAs.

Hfq binding in sRNA–mRNA pairs

A key function of Hfq is to facilitate sRNA–mRNA duplex formation (Møller *et al*, 2002; Zhang *et al*, 2002; Kawamoto *et al*, 2006; Fender *et al*, 2010); this activity seems to require Hfq binding in mRNAs proximal to the site of sRNA pairing, as suggested by studies of *rpoS* mRNA which is regulated by multiple sRNAs (Soper *et al*, 2011). The simultaneous binding of both the sRNA and cognate mRNA by an Hfq hexamer may then accelerate RNA duplex formation at the rim of the protein (Panja *et al*, 2013). To understand where Hfq needs to bind within its ligand to facilitate RNA duplex formation, we performed a meta-gene analysis of Hfq peaks that mapped close to seed pairing regions in known sRNA–mRNA target pairs. In mRNAs, Hfq peaks were significantly more likely to occur 5′ of the respective sRNA interaction site ($P < 0.05$, two-tailed sign test, $n = 17$) (Fig 4A). By contrast, Hfq peaks in sRNAs were found significantly more often 3′ of sRNA seed sequences ($P < 10^{-4}$, two-tailed sign test, $n = 24$) (Fig 4A). This result supports a model whereby Hfq is sandwiched between the mRNA and sRNA of a cognate pair prior to RNA duplex formation (Fig 4B).

The presence of an Hfq site close to an sRNA site in an mRNA improves target regulation (Beisel *et al*, 2012). Therefore, we asked whether our Hfq-binding data could increase the success of sRNA-target predictions. To this end, the top 20 mRNA targets predicted by the CopraRNA algorithm (Wright *et al*, 2013) for each of 17 selected sRNAs were intersected with the list of crosslinked mRNAs, giving 48 predicted mRNA targets with at least one Hfq peak (Fig 4C, Table EV3). Strikingly, inclusion of the Hfq peaks increased the fraction of true positives from 15% to 40% ($P < 10^{-5}$, Fisher's exact test) (Fig 4C).

For experimental validation, we selected the *mglB* mRNA as a new candidate target of Spot42 sRNA. Recognition would occur by a previously established seed sequence within Spot42 (Beisel & Storz, 2011) at a conserved site downstream of the Hfq peak in *mglB* (Figs 4D and EV2). Of note, the levels of MglB, a CRP-cAMP-activated galactose ABC transporter (Zheng *et al*, 2004), are increased in Hfq-deficient cells, predicting that Spot42 represses the *mglB* mRNA in an Hfq-dependent manner (Fig EV2; Sittka *et al*, 2007; Beisel & Storz, 2011). In agreement with this prediction, deletion of *spf* (encoding Spot42) resulted in elevated levels of the *mglB* mRNA (Fig 4E). Reciprocally, we observed a 10-fold repression of this target after pulse-expression of Spot42 (Fig 4F). Spot42 repressed a constitutively transcribed translational *mglB-gfp* fusion, but not a *lacZ-gfp* control, confirming that the regulation occurs at the post-transcriptional level (Fig 4G). To test whether the observed regulation indeed relies on the predicted basepairing, we introduced disruptive mutations in the *mglB-gfp* and Spot42 plasmids (Fig 4H). Deletion of *spf* on the chromosome leads to increased expression of wild-type *mglB-gfp* but not of the mutant *mglB*-gfp* construct (Fig 4H). Likewise, while wild-type Spot42 repressed *mglB-gfp* but not *mglB*-gfp*, the Spot42* mutant repressed *mglB*-gfp* but not *mglB-gfp* (Fig 4H), strongly indicating that the observed regulation indeed relies on basepairing between Spot42 and the *mglB* mRNA, as predicted. In conclusion, these results indicate that knowing which mRNAs are bound by Hfq can dramatically improve the prediction of sRNA targets.

Transcriptome-wide mapping of CsrA-binding sites

Following the successful identification of Hfq-binding sites, we applied our CLIP-seq protocol to CsrA, an RBP that recognizes transcripts very differently compared to Hfq. CsrA has affinity for GGA sequences present in loop regions of hairpins in mRNA 5′UTRs and in a few sRNAs (Vakulskas *et al*, 2015). A *Salmonella* strain carrying a chromosomal *csrA::3xflag* allele was subjected to the same crosslinking and immunoprecipitation strategy described above. As with Hfq, radioactively labeled CsrA-RNA complexes were detected only in crosslinked samples (Fig EV3). Plotting all CsrA peaks obtained from three biological replicates along the *Salmonella* transcriptome revealed a strong enrichment within CsrB and CsrC; almost 40% of reads from all peaks map to these sRNA antagonists of CsrA (Fig 5A and Table EV4), consistent with them being the major cellular ligands of CsrA (Romeo *et al*, 2013). The *glgC* mRNA, the first transcript shown to be directly regulated by CsrA in *E. coli* (Liu *et al*, 1995; Baker *et al*, 2002), was also highly recovered in our experiments (0.5% of reads, Fig 5A and Table EV4).

The CsrB RNA carries multiple hairpins with GGA sequences which serve as high-affinity-binding sites for CsrA. Intriguingly, the read distribution within CsrB is not uniform. Regions with high read densities are separated by low-read regions (Fig 5B). Aligning the CsrA reads on the predicted secondary structure of CsrB, we find that read coverage is highest in the hairpin structures, indicating that these are indeed preferentially bound by CsrA (Fig 5B). Some hairpins show higher coverage than others, perhaps reflecting a hierarchy in CsrA capture by CsrB similar to the proposed step-wise sequestration of the homologous RsmE protein by RsmZ RNA in *Pseudomonas* (Duss *et al*, 2014b). Regarding CsrA mRNA interactions, reads from the *glgC* transcript almost perfectly overlapped

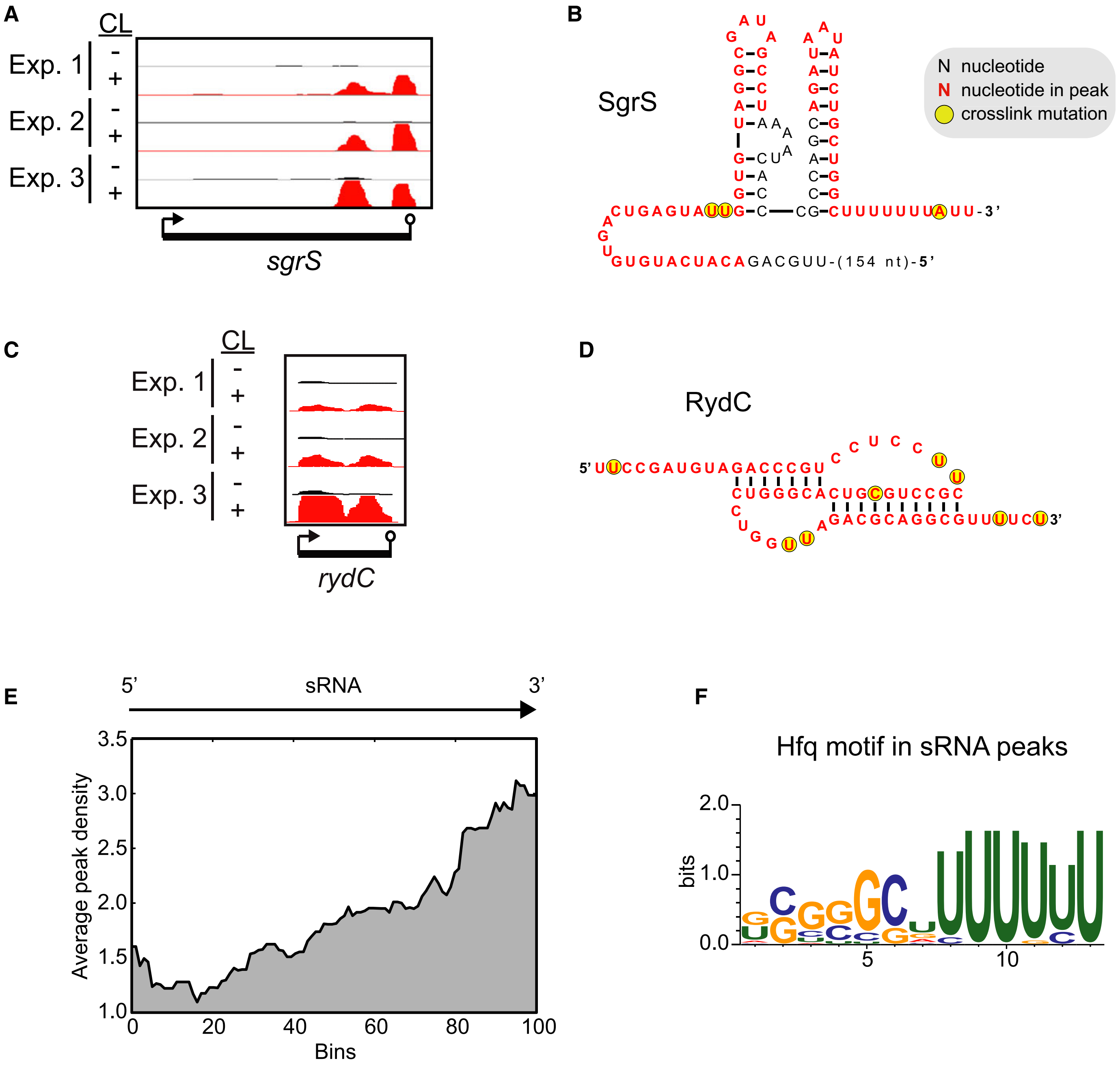


Figure 3. Hfq binding in *Salmonella* sRNAs.

A Read coverage in libraries from crosslinked and non-crosslinked samples at the *sgrS* locus. CL: crosslinking

B Predicted secondary structure of the sRNA SgrS. Nucleotides corresponding to a Hfq peak and positions of crosslink-induced mutations are color coded as highlighted in the legend.

C Read coverage in libraries from crosslinked and non-crosslinked samples at the *rydC* locus. CL: crosslinking.

D Predicted pseudoknot structure of the sRNA RydC. Nucleotides corresponding to an Hfq peak and positions of crosslink-induced mutations are color coded as highlighted in (B).

E Meta-gene analysis of the peak distribution along *Salmonella* sRNAs. Length normalization was achieved through proportional binning according to the different lengths of the sRNA sequences.

F Consensus motif generated by MEME using sequences of peaks mapping to sRNAs as input.

with a GGA-containing hairpin structure in the *glgC* leader (Fig 5C), which was previously defined as the element through which CsrA exercises translational repression in *E. coli* (Baker *et al*, 2002). The detection of CsrA peaks in these two well-documented targets of CsrA suggests that our method readily captures *bona fide* CsrA-binding sites (Fig 5A–C).

Published online: April 4, 2016

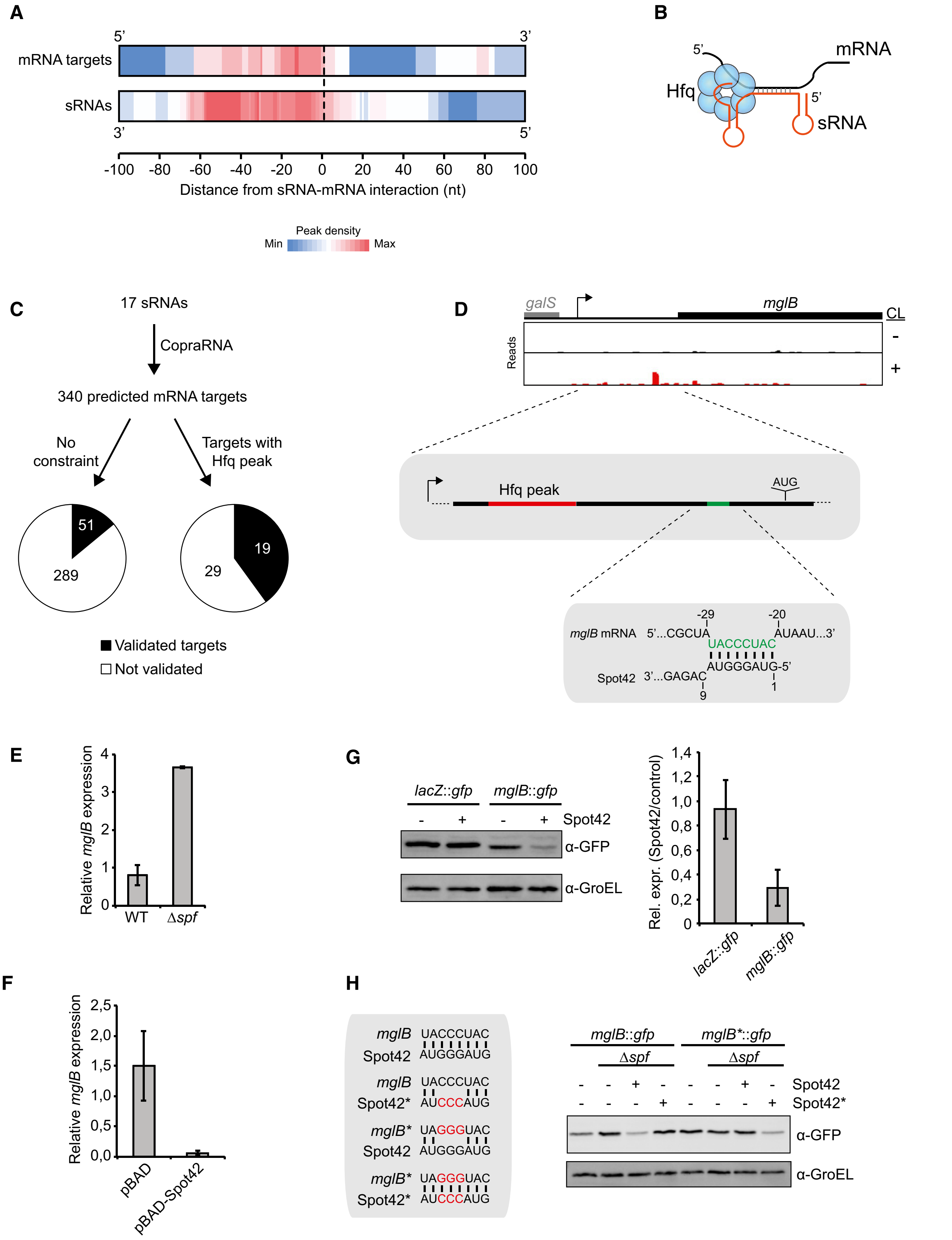


Figure 4.

Published online: April 4, 2016

◀ **Figure 4. Hfq binding in validated sRNA–mRNA pairs.**

A Distribution of Hfq peaks with respect to sRNA interaction sites in mRNA targets and seed sequences in sRNAs, respectively.

B Putative model of Hfq interaction with cognate sRNA–mRNA pairs.

C Workflow for the integration of Hfq peak information during sRNA-target prediction using CopraRNA. The pie charts show the number of previously validated targets among all predictions, or among predicted targets with Hfq peaks, respectively.

D Read coverage from Hfq CLIP-seq at the *mglB* locus (top), location of the detected Hfq peak (red) and the predicted Spot42 interaction site (green) in the *mglB* 5′UTR (middle), and the predicted basepair interaction between Spot42 and *mglB* (bottom). The Spot42 interaction site in *mglB* is highlighted in green.

E qRT–PCR analysis of *mglB* mRNA expression in wt *Salmonella* or in an isogenic Δ*spf* strain. Samples were collected from cells grown in LB medium to an optical density of 0.3 (OD_{600}). Means and error bars representing standard deviations are based on two biological replicates.

F qRT–PCR analysis of *mglB* mRNA expression in *Salmonella* Δ*spf* 10 min after induction of Spot42 overexpression from plasmid pBAD–Spot42. Plasmid pBAD was used as a control. Means and error bars representing standard deviations are based on two biological replicates.

G Western blot analysis of GFP expression from plasmid-expressed translational *lacZ-gfp* and *mglB-gfp* fusions in the presence or absence of Spot42 overexpression. Quantification of Western blot signals is shown on the right. Means and error bars representing standard deviations are based on three biological replicates. GFP fusion proteins were detected with an anti-GFP antibody, while an anti-GroEL antibody was used to determine the amount of protein loaded on the gel.

H Western blot analysis of GFP expression from the wild-type *mglB-gfp* or mutant *mglB**-*gfp* fusions upon deletion and overexpression of wild-type Spot42 or the Spot42* mutant. The predicted interactions between Spot42 and *mglB*, as well as the introduced mutations, are shown.

Source data are available online for this figure.

CsrA consensus motif

We called a total of 467 CsrA peaks, most of which map to within mRNAs (Fig 6A and Table EV4). Meta-gene analysis showed an enrichment of peaks in 5′UTRs compared to CDSs and 3′UTRs, with the strongest enrichment of peaks close to start codons, consistent with CsrA being a regulator of translation initiation (Fig 6B).

High-affinity CsrA–RNA interactions are defined by both RNA sequence and structure (Romeo *et al*, 2013). Interrogation of the CsrA peaks showed that each contained at least one minimal GGA triplet and more than half of them an ANGGA sequence (Fig 6C). Searching all peak regions using the MEME algorithm, we established [A/C]UGGA as the CsrA recognition motif in *Salmonella* (Fig 6D).

Similar to Hfq, we observed that crosslinking of CsrA to RNA frequently causes mutations during reverse transcription. T to C transitions were most prominent (Fig 6E, Table EV5), and these were most often found immediately upstream of a GGA motif (Fig 6E). To analyze the structural context of CsrA-binding sites, we performed CMfinder analysis on all CsrA peaks (Yao *et al*, 2006). Two of the resulting motifs, the one with the highest rank score and the one detected in the most peak sequences (Fig 6F left and right, respectively), consist of stem-loops with a GGA sequence present in the loop regions. Thus, our CLIP analysis confirms the preference for CsrA to interact with AUGGA sequences present in apical loops of hairpin structures. These are the first global data to prove the previous biochemical and genetical studies of individual CsrA ligands, which increasingly suggested ANGGA as a general recognition motif in a variety of bacterial species (Valverde *et al*, 2004; Dubey *et al*, 2005; Majdalani *et al*, 2005; Mercante *et al*, 2006; Babitzke *et al*, 2009; Lapouge *et al*, 2013).

CsrA regulates *Salmonella* virulence genes

Binding of CsrA to target mRNAs typically results in reduced mRNA translation and/or stability (Romeo *et al*, 2013). Since the vast majority of the CsrA sites detected here were previously unknown, we wondered whether they were functional in terms of CsrA-mediated gene regulation. One primary genomic area of CsrA peak density was the invasion gene island SPI-1; likewise, a KEGG pathway analysis suggested enrichment of CsrA peaks in mRNAs encoding *Salmonella* virulence proteins (Fig 7A and B). Our crosslinking data (Table EV4) not only support the previously proposed direct regulation of *hilD* mRNA (encoding a SPI-1 transcription factor) by CsrA (Martinez *et al*, 2011), but also predict CsrA to target dozens of additional virulence-associated mRNAs from both *Salmonella*'s pathogenicity islands and the core genome (Appendix Fig S2).

To test whether the presence of CsrA peaks correlates with CsrA-mediated gene regulation, we constructed translational *gfp*-fusion reporters (Corcoran *et al*, 2012) to several virulence-associated ORFs from the core genome (*sopD2*) or the SPI-1 locus (*sic-sip* and *prg* operons). GFP fusion plasmids were transformed into Δ*csrB*Δ*csrC* cells harboring either a plasmid expressing CsrB, or an empty control plasmid, reasoning that CsrB-mediated titration of CsrA will translate into GFP reporter regulation. This strategy was chosen to circumvent the genetic instability observed in *csrA* deletion strains (Altier *et al*, 2000). While co-expression of CsrB had no effect on a *lacZ-gfp* control plasmid (pXG10-SF), it caused a strong derepression of a *glgC-gfp* fusion chosen as positive control (Fig EV4), arguing that this experimental setup faithfully monitors CsrA-mediated regulation.

SopD2 is an effector protein that promotes *Salmonella* replication inside macrophages (Figueira *et al*, 2013), and CLIP-seq data identified several CsrA peaks in the *sopD2* 5′UTR and CDS (Fig 7C). Western blot analysis showed that *sopD2-gfp* expression is repressed when CsrA activity is increased as a result of deletion of *csrB* and *csrC* (Fig 7D). This is reversed by complementing the double sRNA deletion strain with *csrB* on a plasmid (Fig 7D). A CsrA peak in the 5′UTR of *sopD2* overlaps with a predicted RNA hairpin structure with two GGA motifs in the loop (Fig 7E). A *sopD2-gfp* fusion in which both GGA motifs were each replaced by CCU totally abolished the regulation, strongly indicating that CsrA directly represses the production of SopD2 (Fig 7E). In further support of this, overexpression of CsrB upregulates the synthesis of endogenous SopD2 in wild-type *Salmonella* (Fig EV5).

The *prgHIJK-orgA* operon encodes components of the SPI-1 type III secretion system needed for host cell invasion, and CsrA peaks were detected in its four-first cistrons (Fig 7F). Western blot analysis with translational fusions encompassing cistron junctions with the downstream cistron being fused to *gfp* showed that translation of *prgI* and *prgJ* is activated upon CsrB overexpression, whereas

Published online: April 4, 2016

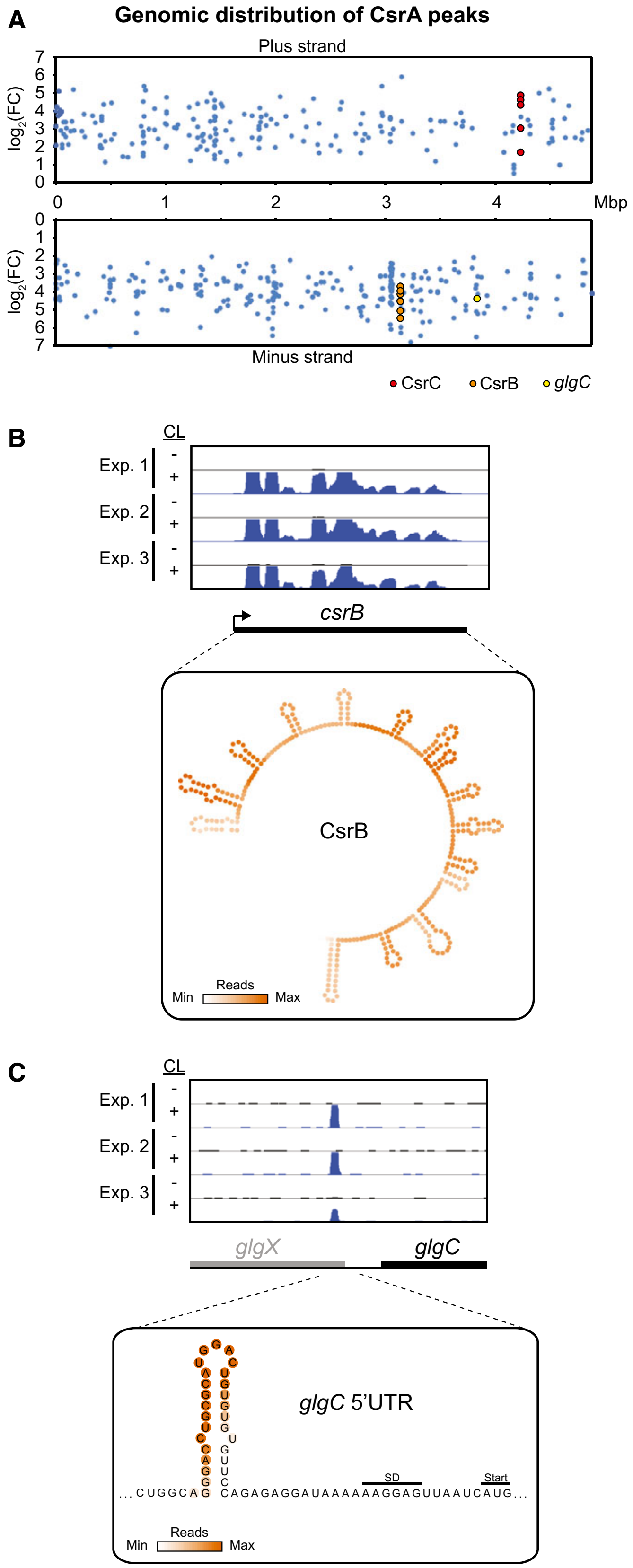


Figure 5. CLIP-seq of *Salmonella* CsrA-3xFLAG captures previously known CsrA-binding sites.

A Fold change (*y*-axis) and genomic position (*x*-axis) of CsrA peaks. Peaks mapping to the known CsrA ligands CsrB, CsrC, and *glgC* are indicated.

B Read coverage from CsrA CLIP-seq at the *csrB* locus (top). A heat map of the average read coverage at the *csrB* locus superimposed on the predicted secondary structure of *Salmonella* CsrB (bottom). The CsrB structure was predicted by MFOLD (Zuker, 2003).

C Read coverage from CsrA CLIP-seq at the *glgC* locus (top). A heat map of the average read coverage at the *glgC* locus superimposed on the predicted secondary structure of the 5′UTR of the *Salmonella glgC* mRNA (bottom).

prgK is not affected (Fig 7G). Of note, the major peaks are located in *prgI* and *prgJ* (Fig 7F). Similarly, the *sicA-sipBCDA-iacP* operon encodes a protein chaperone (SicA), four effector proteins (SipB, SipC, SipD, and SipA), and a putative acyl carrier protein (IacP), and CsrA peaks are distributed across this operon (Fig 7H). Of the four fusions cloned from this operon, three (*sicA, sipC, and sipA*) were clearly upregulated upon CsrB overexpression, indicating that expression from the respective cistrons is repressed by CsrA (Fig 7I). In conclusion, the results shown in Fig 7 strongly indicate that CsrA peaks indeed mark mRNAs that are under direct control of CsrA and suggest that direct regulation of virulence functions by CsrA includes many more mRNAs than previously known.

Discussion

Historically, molecular biologists have focused on the interactions between individual proteins with target nucleic acids *in vitro*, but this approach does not scale well and fails to account for the complexity observed in transcriptional networks. Post-genomic approaches can now potentially provide the global data required to understand post-transcriptional gene regulation in bacteria (Barquist & Vogel, 2015). Specifically, *in vivo* crosslinking methods can determine protein-binding sites within RNA at high resolution and permit stringent purification that diminishes non-specific contamination. Nevertheless, these CLIP-seq approaches have been associated with considerable background noise that, if left uncorrected, increased the identification of false positive interactions (Friedersdorf & Keene, 2014). Here, we have sequenced libraries prepared from both UV crosslinked and non-crosslinked bacterial cultures to control for background RNA, yielding a high-confidence transcriptome-wide map of the binding sites of the two global RNA-binding proteins Hfq and CsrA.

We have shown that Hfq selectively and primarily crosslinks to *Salmonella* mRNAs and sRNAs (Fig 2), in accordance with our previous Hfq coIP results (Sittka *et al*, 2008; Chao *et al*, 2012). More importantly, while relatively few Hfq–sRNA interactions have been studied in biochemical or structural detail, we can faithfully reproduce such results with single-nucleotide resolution in our crosslinking experiment, as shown in Fig 3 for the model sRNAs RydC and SgrS (Ishikawa *et al*, 2012; Dimastrogiovanni *et al*, 2014). Global analysis revealed that Hfq peaks in mRNAs are enriched in 5′UTRs and 3′UTRs as compared to CDS regions (Fig 2), consistent with a role for Hfq in both sRNA-dependent regulation at mRNA 5′ regions and 3′ end-dependent processes. Analysis of Hfq peak density over the *Salmonella* transcriptome revealed strong enrichment in transcripts expressed from the major pathogenicity islands SPI-1 and SPI-2 (Fig 2B). This may in part be explained by the higher content of A and U residues in these transcripts compared

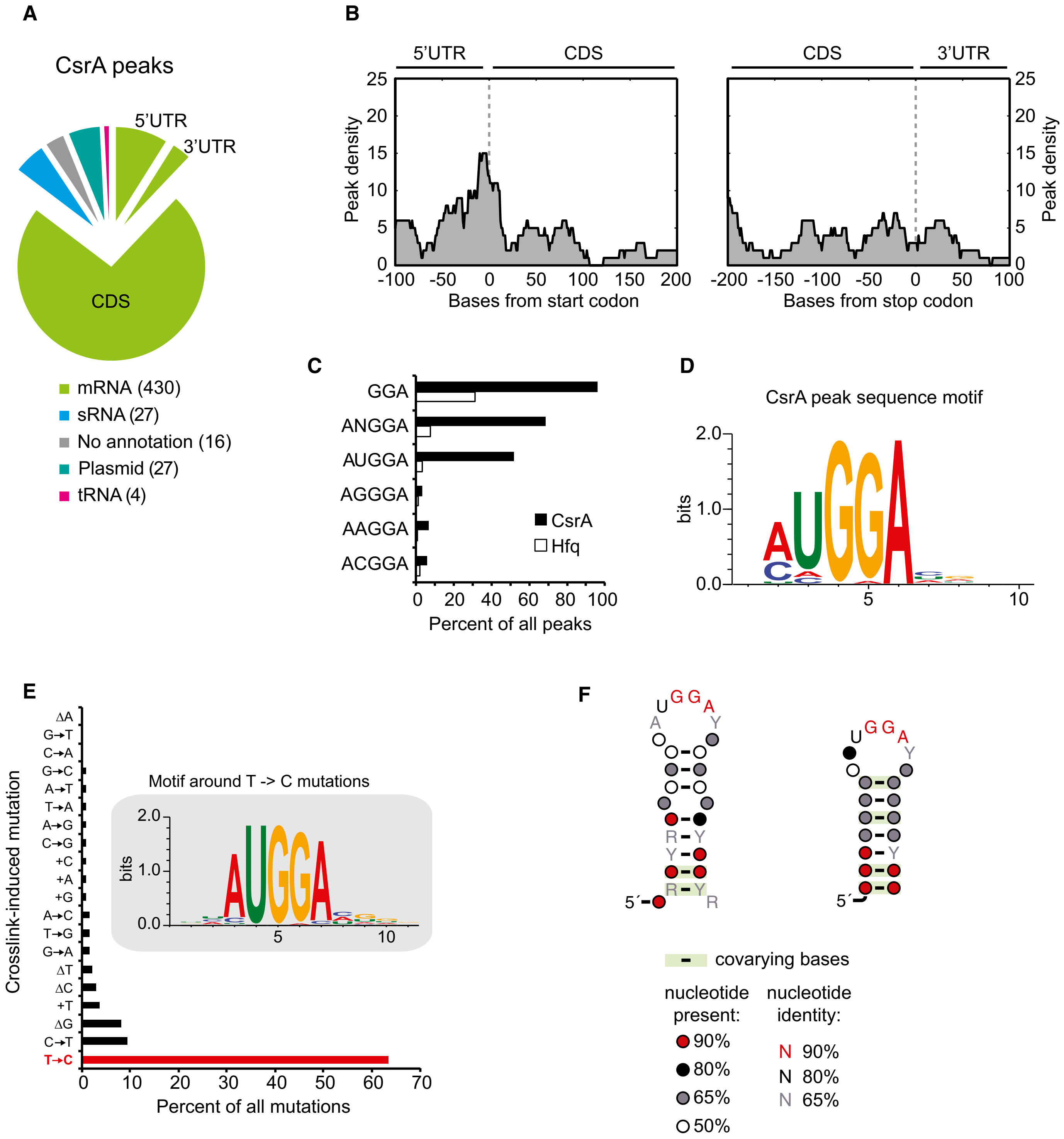


Figure 6. Sequence and structure analysis of CsrA-binding sites.

A Distribution of CsrA peaks among the indicated RNA classes. Numbers in parenthesis represent the number of called peaks that were mapped within annotations belonging to the respective RNA class.

B Meta-gene analysis of CsrA peaks around start and stop codons. For this analysis, only those start and stop codons were used that are flanked by a 5′UTR or 3′UTR, respectively.

C Percentage of peaks that contain the indicated sequences.

D Consensus motif generated by MEME based on all CsrA peak sequences.

E Percentage of the occurrence of the indicated mutations among all crosslink-specific mutations found within CsrA peaks. The inset shows the consensus motif generated with MEME using sequences flanking a crosslink-specific T to C mutation as input.

F Consensus motifs generated by CMfinder based on all CsrA peaks.

Published online: April 4, 2016

Figure 7. CsrA plays a major role in the regulation of *Salmonella* virulence genes.

A CsrA peak density distribution along the *Salmonella* chromosome in bins of 2 × 10^4 basepairs. The genomic positions of *Salmonella* pathogenicity islands SPI-1 and SPI-2 are indicated.

B KEGG pathways that were found significantly enriched among gene annotations to which CsrA peaks were mapped. Pathways that are related to *Salmonella* pathogenicity are highlighted in red.

C Read coverage from CsrA CLIP-seq at the *sopD2* locus. Light blue bars represent called peaks.

D Western blot analysis of SopD2-GFP expression from a translational *sopD2-gfp* fusion on a plasmid in the indicated strain backgrounds. Plus sign indicates the presence of plasmid pCsrB. Minus sign indicates the presence of the control vector pJV300. SopD2-GFP signals were detected with an anti-GFP antibody. Expression of GroEL served as a loading control and was detected with an anti-GroEL antibody.

E Predicted secondary structure of the *sopD2* 5′UTR. Peak position, GGA motifs, and introduced mutations are indicated. GFP fluorescence measurements from the wild-type *sopD2-gfp* fusion or a 2xCCU mutant upon *csrBcsrC* deletion and CsrB complementation. Means and error bars representing standard deviations are based on three independent experiments.

F Read coverage at the *prgHIJK-orgAB* locus from a CsrA CLIP-seq experiment.

G Western blot analysis of the expression from the indicated plasmid-borne translational GFP fusions in the presence of plasmids pCsrB (plus signs) or pJV300 (minus signs).

H Read coverage at the *sicA-sipBCDA-iacP* locus from a CsrA CLIP-seq experiment.

I Western blot analysis of the expression from the indicated plasmid-borne translational GFP fusions in the presence of plasmids pCsrB (plus signs) or pJV300 (minus signs).

Source data are available online for this figure.

to those expressed from the core genome (Hensel, 2004). Comprehensive analysis of sRNA peaks revealed a strong enrichment of Hfq binding at 3′ ends (Fig 3). The highly enriched consensus motifs found in peak sequences from either mRNA 3′UTRs or sRNAs, respectively, both resemble the 3′ region of Rho-independent terminators (Figs 2, 3 and EV1) and were indeed found in 3′UTRs of mRNAs predicted to transcriptionally terminate in a Rho-independent manner (Fig 2).

The strong evidence for Hfq binding to 3′ ends in mRNAs and sRNAs presented here agrees with previous reports on individual Hfq ligands. Hfq protects RNA from 3′ to 5′ exonuclease activity by binding to, and stimulating the addition of, non-templated poly(A) sequences to RNA 3′ ends by poly(A) polymerase PAPI (Hajnsdorf & Regnier, 2000; Le Derout *et al*, 2003). The sRNA SgrS strongly depends on Hfq binding at its 3′ poly(U) tail for both stability and target regulation (Otaka *et al*, 2011), and the destabilization of SgrS in the absence of Hfq is dependent on the exonuclease PNPase (Andrade *et al*, 2012).

That Hfq binds so commonly to mRNA 3′ ends may be very relevant for sRNA evolution. Cloning or RNA-seq-based studies have identified many sRNAs derived from mRNA 3′UTRs (Vogel *et al*, 2003; Kawano *et al*, 2005; Sittka *et al*, 2008; Chao *et al*, 2012). Whether these sRNAs are produced from internal promoters or by endonucleolytic cleavage of the parental mRNA, they often possess a Rho-independent terminator shared with the mRNA expressed from the same locus (Miyakoshi *et al*, 2015b). Several 3′ UTR-derived sRNAs have been shown to be functional, for example DapZ (Chao *et al*, 2012), MicL (Guo *et al*, 2014), or SroC (Miyakoshi *et al*, 2015a), suggesting that mRNA 3′UTRs may serve as evolutionary birthplaces for sRNAs (Miyakoshi *et al*, 2015b; Updegrove *et al*, 2015). This extends to other types of regulatory transcripts such as recently discovered sRNA sponges that are made from the 3′ end of tRNA precursors (Lalaouna *et al*, 2015).

A key finding from our analysis of the crosslinking data is that we were able to locate Hfq-binding sites in relation to sRNA–mRNA interaction sites (Fig 4). Our observation of preferential binding of Hfq to 5′ of the sRNA interaction site in an mRNA target, and 3′ of the seed sequence in the recognizing sRNA, supports a model whereby Hfq brings the two RNAs together to facilitate RNA duplexing. We used this global information on Hfq binding to substantially improve sRNA-target predictions (Fig 4), illustrating how global RNA–protein interaction maps can foster a better understanding of post-transcriptional networks and discovering the *mglB* mRNA as a target for the sRNA Spot42 (Fig 4). MglB is a transporter of the non-preferred carbon source galactose, and its expression is activated by CRP–cAMP (Zheng *et al*, 2004). Thus, the regulation of *mglB* by Spot42 fits with a proposed model in which Spot42 and CRP form a feed-forward loop to reduce leaky expression of proteins during carbon foraging (Fig EV2; Beisel & Storz, 2011).

The fact that Hfq binds RNA on three distinct faces of the hexamer, each with a different sequence preference, produces a challenge for CLIP-seq methods in that ligation of sequencing adapters to RBP-bound RNA, as well as UV irradiation, may introduce biases in binding site detection. This may explain why our Hfq CLIP-seq data contrast with a recent crosslinking study of Hfq in *E. coli* (Tree *et al*, 2014). This latter study identified neither the 3′-located terminator-like consensus motif nor an enrichment of Hfq-binding sites in sRNA 3′ ends. Instead, the authors concluded that Hfq binding occurs in the seed sequences located in the middle or at the 5′ end of sRNAs. These differences can be explained by differences in the protocols: 3′ adapter ligation to RNA in complex with Hfq (Tree *et al*, 2014) versus adapter ligation after the RNA fragments are released from Hfq (this study). As RNA 3′ ends may not be accessible to ligation when bound to the proximal side of Hfq, adapter ligation to purified RNA as performed here may be the preferred strategy for CLIP approaches when studying proteins that target RNA 3′ ends.

In addition, Tree *et al* (2014) reported a general ARN motif in Hfq crosslink regions, which seemed consistent with structural data on the interaction between the distal face of Hfq and A-rich sequences (Link *et al*, 2009), and the involvement of mRNA located ARN sequences in sRNA-dependent regulation (Salim & Feig, 2010; Beisel *et al*, 2012; Salim *et al*, 2012; Peng *et al*, 2014). Reviewing our CLIP-seq data, on the one hand, almost all (38/39) Hfq peaks in mRNAs known to be targeted by sRNAs (including *rpoS, ompA, ompC, cfa,* and *mglB*) contain at least one ARN motif (Table EV1). On the other hand, we only detected Hfq peaks in 30% of the previously described sRNA targets (Table EV1) (Wright *et al*, 2013), and we did not observe a significant enrichment of ARN motifs among the mRNA peak sequences compared to randomly selected

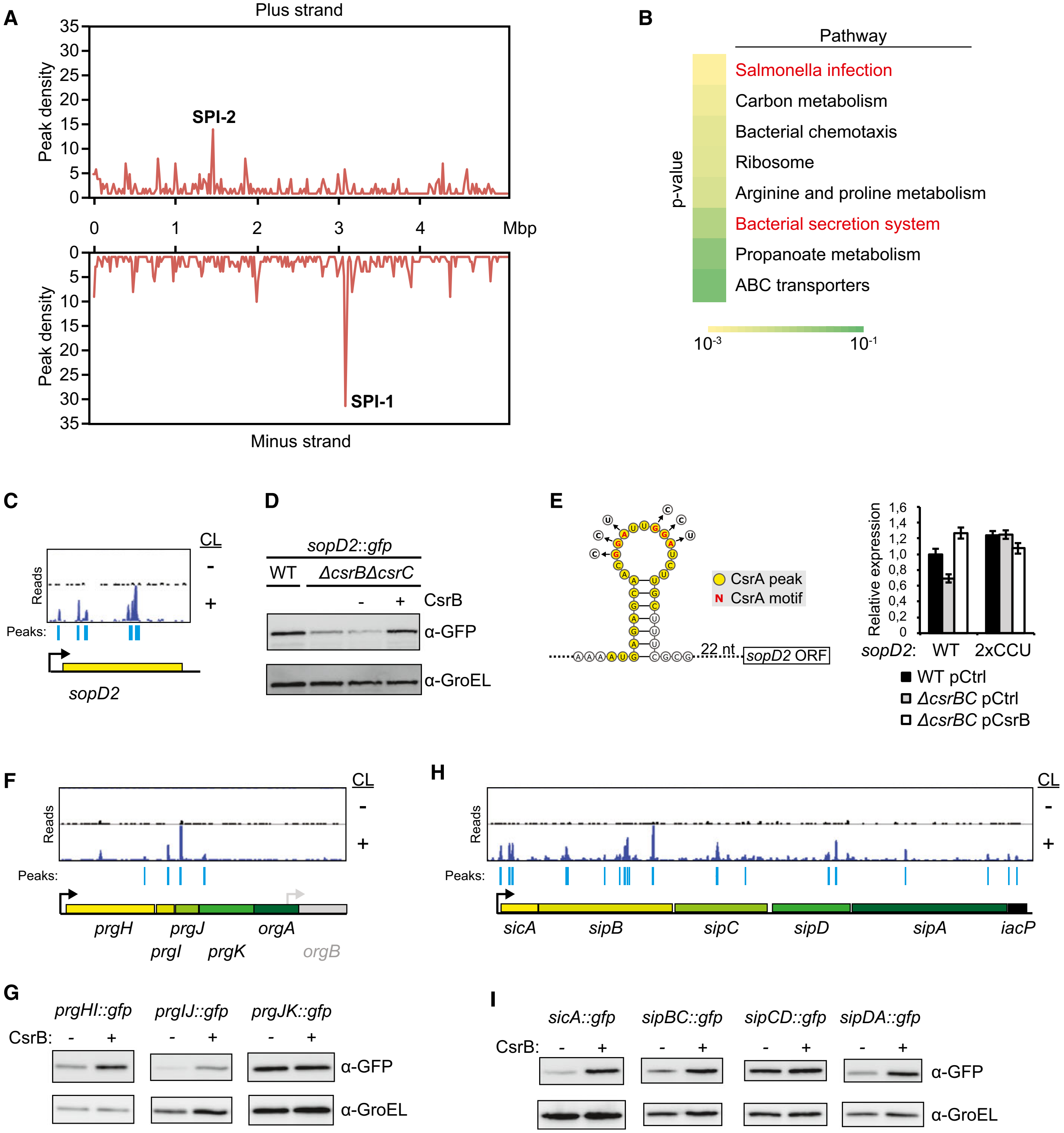


Figure 7.

sequences. One explanation for this discrepancy may be that uridines are more prone to crosslink than other nucleosides (Sugimoto *et al*, 2012); this bias together with the above-discussed adaptor ligation issues may explain why we preferentially detect binding of Hfq at 3′-located U-rich sequences, while the different adapter ligation strategy forced preferential detection of A-rich sequences in the previous *E. coli* study (Tree *et al*, 2014).

Moreover, the canonical view that sRNAs generally interact with the proximal side of Hfq and mRNA targets with the distal side has already been challenged: a recent study showed that some sRNAs use ARN sequences to interact with the distal side of Hfq, whereas their cognate targets harbor 5′UTR-located UA-rich rim-binding sequences (Schu *et al*, 2015). In support of this finding, we find crosslinking mutations in an ARN sequence in the sRNA ChiX and in a UA-rich

Published online: April 4, 2016

sequence in the cognate target mRNA *chiP* (*ybfM*) (Table EV2). Taken together, we propose that mapping of the *in vivo* binding events at each of the three Hfq interaction faces, applying CLIP-seq to mutant Hfq proteins, should be undertaken to further test the current model of distinct "sRNA" and "mRNA" binding faces of Hfq.

These issues with Hfq notwithstanding, the successful application of our crosslinking protocol to CsrA, an RBP with very different targets and recognition mode to Hfq, strongly supports the general applicability of our crosslinking protocol. In contrast to Hfq-binding regions, the vast majority of the detected CsrA-binding sites contain the crucial GGA motif for CsrA–RNA interactions (Figs 5 and 6; Vakulskas *et al*, 2015). CsrA is known to regulate virulence gene expression in *Salmonella*, and a direct interaction between CsrA and *hilD* mRNA, encoding a transcriptional activator of SPI-1, has been described (Martinez *et al*, 2011). In addition to binding *hilD* mRNA, our crosslinking data suggests that CsrA binds to a plethora of virulence-associated mRNAs (Appendix Fig S2). The regulatory potential of newly discovered CsrA-binding sites in virulence-associated mRNAs was confirmed using GFP reporters (Fig 7), consistent with previous reports showing that the levels of some of these mRNAs depend on the intracellular CsrA concentration (Altier *et al*, 2000; Lawhon *et al*, 2003). Even though our validation of CsrA targets is far from comprehensive, it already expands the number of *Salmonella* virulence mRNAs that are post-transcriptionally regulated by CsrA sixfold. Based on our findings, it is likely that more virulence mRNAs are directly regulated by CsrA.

In *Escherichia coli*, the Hfq-dependent McaS sRNA was recently reported to titrate CsrA, suggesting that sRNAs other than CsrB and CsrC may be functional CsrA interaction partners (Jørgensen *et al*, 2013). Interestingly, we also detected binding sites for CsrA in sixteen sRNAs in addition to CsrB and CsrC (Fig 6 and Table EV4), although the read coverage of these additional sRNAs was far below that of CsrB and CsrC. The majority of these sRNAs (14 of 16) carry between one and six GGA motifs, and many of the corresponding peak sequences (12 of 16) fold into hairpins with GGA sequences in the loops (Appendix Fig S3), suggesting that they possess bona fide CsrA-binding sites. Apart from a few well-characterized Hfq-binding sRNAs, of which only one (SdsR) harbors GGA motifs, the majority of the sRNAs that crosslinked to CsrA are uncharacterized. Comparative expression analysis revealed that several of these sRNAs (STnc1890, STnc2080, STnc1210, STnc1480, PinT, and SdsR) are induced in late stationary phase, a growth condition in which CsrB and CsrC are repressed (Kröger *et al*, 2013). This suggests that these six sRNAs may compete with CsrB and CsrC under specific conditions. Future studies will be required to determine whether or not these sRNAs are functional CsrA antagonists, or perhaps are regulated by CsrA.

Bacteria express a plethora of regulatory RBPs for which no global binding site information is available. Examples of these include proteins with RNA-binding domains found in cold-shock proteins (the Csp family of proteins) and proteins such as ProQ that possess a FinO-like RNA-binding domain (Phadtare *et al*, 1999; Mark Glover *et al*, 2015). We believe that our procedure for global mapping of the Hfq and CsrA interactomes with cellular RNA will lay the foundations for future studies of other important bacterial RBPs and may also rapidly identify proteins with putative RNA-binding potential. Such studies should be a major future direction in the study of post-transcriptional phenomena in bacteria and will shed light on this shadowy area of gene regulation.

Materials and Methods

Oligodeoxyribonucleotides

DNA oligonucleotides are listed in Appendix Table S1.

Bacterial strains and plasmids

All experiments were performed with *Salmonella enterica* serovar Typhimurium strain SL1344 or derivatives thereof as listed in Appendix Table S2. All plasmids used in this study are listed in Appendix Table S3. Construction of strains and plasmids is described in Appendix Supplementary Methods. The addition of a FLAG-tag to Hfq or CsrA affected neither bacterial growth nor regulation of known Hfq or CsrA targets, indicating that the tag did not compromise protein function (Appendix Fig S4).

UV crosslinking, immunoprecipitation, and RNA purification

For each biological replicate, 200 ml bacterial culture was grown until an OD_{600} of 2.0. Half of the culture was directly placed in a 22 × 22 cm plastic tray and irradiated with UV-C light at 800 mJ/cm^2. Cells were pelleted in 50 ml fractions by centrifugation for 40 min at 6,000 *g* and 4°C, resuspended in 800 μl NP-T buffer (50 mM NaH_2PO_4, 300 mM NaCl, 0.05% Tween, pH 8.0) and mixed with 1 ml glass beads (0.1 mm radius). Cells were lysed by shaking at 30 Hz for 10 min and centrifuged for 15 min at 16,000 *g* and 4°C. Cell lysates were transferred to new tubes and centrifuged for 15 min at 16,000 *g* and 4°C. The cleared lysates were mixed with one volume of NP-T buffer with 8 M urea, incubated for 5 min at 65°C in a thermomixer with shaking at 900 rpm and diluted 10× in ice-cold NP-T buffer. Anti-FLAG magnetic beads (Sigma) were washed three times in NP-T buffer (30 μl 50% bead suspension was used for a lysate from 100 ml bacterial culture), added to the lysate, and the mixture was rotated for one hour at 4°C. Beads were collected by centrifugation at 800 *g*, resuspended in 1 ml NP-T buffer, transferred to new tubes, and washed 2× with high-salt buffer (50 mM NaH_2PO_4, 1 M NaCl, 0.05% Tween, pH 8.0) and 2× with NP-T buffer. Beads were resuspended in 100 μl NP-T buffer containing 1 mM $MgCl_2$ and 2.5 U benzonase nuclease (Sigma) and incubated for 10 min at 37°C in a thermomixer with shaking at 800 rpm, followed by a 2-min incubation on ice. After one wash with high-salt buffer and two washes with CIP buffer (100 mM NaCl, 50 mM Tris–HCl pH 7.4, 10 mM $MgCl_2$), the beads were resuspended in 100 μl CIP buffer with 10 units of calf intestinal alkaline phosphatase (NEB) and incubated for 30 min at 37°C in a thermomixer with shaking at 800 rpm. After one wash with high-salt buffer and two washes with PNK buffer (50 mM Tris–HCl pH 7.4, 10 mM $MgCl_2$, 0.1 mM spermidine), one-tenth of the beads was removed for subsequent Western blot analysis. The remaining beads were resuspended in 100 μl PNK buffer with 10 U of T4 polynucleotide kinase and 10 μCi γ-^{32}P-ATP and incubated for 30 min at 37°C. After three washes with NP-T buffer, the beads were resuspended in 20 μl Protein Loading buffer (0.3 M Tris–HCl pH 6.8, 0.05% bromophenol blue, 10% glycerol, 7% DTT) and incubated for 3 min at 95°C. The magnetic beads were collected on a magnetic separator, and the supernatant was loaded and separated on a 15% SDS–polyacrylamide gel. RNA–protein complexes were transferred

Published online: April 4, 2016

to a nitrocellulose membrane, the protein marker was highlighted with a radioactively labeled marker pen and exposed to a phosphor screen for 30 min. The autoradiogram was used as a template to cut out the labeled RNA–protein complexes from the membrane. Each membrane piece was further cut into smaller pieces, which were incubated for 30 min in a thermomixer at 37°C with shaking at 1,000 rpm in 400 μl PK solution [50 mM Tris–HCl pH 7.4, 75 mM NaCl, 6 mM EDTA, 1% SDS, 10 U of SUPERaseIN (Life Technologies) and 1 mg/ml proteinase K (ThermoScientific)] whereafter 100 μl 9 M urea was added and the incubation was continued for additional 30 min. About 450 μl of the PK solution/urea was mixed with 450 μl phenol:chloroform:isoamyl alcohol in a phase-lock tube and incubated for 5 min in a thermomixer at 30°C with shaking at 1,000 rpm followed by centrifugation for 12 min at 16,000 *g* and 4°C. The aqueous phase was precipitated with 3 volumes of ice-cold ethanol, 1/10 volume of 3 M NaOAc pH 5.2, and 1 μl of GlycoBlue (Life Technologies) in LoBind tubes (Eppendorf). The precipitate was pelleted by centrifugation (30 min, 16,000 *g*, 4°C), washed with 80% ethanol, centrifuged again (15 min, 16,000 *g*, 4°C), dried 2 min at room temperature, and resuspended in 10 μl sterile water.

cDNA library preparation

To enable sequencing on Illumina instruments, libraries were prepared using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (#E7300, New England Biolabs) according to the manufacturer's instructions. About 2.5 μl purified RNA (or sterile water as negative control) was mixed with 0.5 μl 3′ SR Adaptor (diluted 1:10) and 0.5 μl nuclease-free water, incubated for 2 min at 70°C and chilled on ice. After addition of 5 μl 3′ ligation reaction buffer and 1.5 μl 3′ ligation enzyme mix, the samples were incubated for 60 min at 25°C. About 0.25 μl SR RT primer and 2.5 μl nuclease-free water were added followed by incubation for 5 min at 75°C, 15 min at 37°C, and 15 min at 25°C. For ligation of the 5′ adaptor, the sample was mixed with 0.5 μl 5′ SR adaptor (denatured, diluted 1:10), 0.5 μl 10× ligation reaction buffer, and 1.24 μl ligation enzyme mix and incubated for 60 min at 25°C. cDNA synthesis was carried out by the addition of 4 μl first strand synthesis reaction buffer, 0.5 μl murine RNase inhibitor, and 0.5 μl Protoscript reverse transcriptase and incubation at 50°C for 60 min. The reverse transcription activity was inhibited by a 15-min incubation at 70°C. The cDNA was amplified by PCR by mixing 10 μl cDNA sample with 25 μl 2× LongAmp Taq PCR master mix, 1.25 μl SR primer and 17.5 μl nuclease-free water in a thermal cycler with the following program: 30 s at 94°C, 18 rounds of (15 s at 94°C, 30 s at 62°C, and 15 s at 70°C). The PCRs were purified on columns (QIAGEN), eluted in 10 μl sterile water, and loaded on 6% polyacrylamide gels with 7 M urea together with a 50 bp DNA size marker (ThermoScientific). Gels were stained with SYBRGold (Life Technologies), and fragments between 140 and 250 bp were excised from the gels. Elution of DNA fragments was performed in 500 μl DNA elution buffer (NEB) at 16°C overnight in a thermomixer at 1,000 rpm followed by EtOH precipitation. Pellets were resuspended in 10 μl sterile water. About 2 μl gel-purified DNA was mixed with 25 μl 2× LongAmp Taq PCR master mix, 2 μl each of primer JVO-11007 and JVO-11008 (10 μM), and 19 μl sterile water and amplified using the following program: 30 s at 94°C, 6 rounds of (15 s at 94°C, 30 s at 60°C, and 15 s at 65°C). PCRs were purified on columns (QIAGEN) and eluted in 15 μl sterile water.

Sequencing

High-throughput sequencing was performed at vertis Biotechnologie AG, Freising, Germany. Twelve cDNA libraries were pooled on an Illumina NextSeq 500 mid-output flow cell and sequenced in paired-end mode (2 × 75 cycles). Raw sequencing reads in FASTQ format and coverage files normalized by DESeq2 size factors are available via Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under accession number GSE74425.

Processing of sequence reads and mapping

To assure high sequence quality, read 1 (R1) and read 2 (R2) files containing the Illumina paired-end reads in FASTQ format were trimmed independently from each other with a Phred score cutoff of 20 by the program fastq_quality_trimmer from FASTX toolkit version 0.0.13. In the same step, after quality trimming NEB, R1 and R2 3′-adapters (R1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC, R2: GATCGTCGGACTGTAGAACTCTGAACGTGTAGATCTCGGTGGTCGCCGTATCATT) were trimmed using Cutadapt version 1.7.1 (Martin, 2011) and reads without any remaining bases were discarded. Afterward, reads without a mate in the complementary read file were excluded using cmpfastq (http://compbio.brc.iop.kcl.ac.uk/software/cmpfastq.php). In order to remove putative PCR duplicates, paired-end reads were collapsed using FastUniq (Xu *et al*, 2012). Subsequently, a size filtering step was applied in which read pairs with at least one read shorter than 12 nt or longer than 25 nt were eliminated. The collections of remaining reads were mapped to the *Salmonella* Typhimurium SL1344 chromosome (NCBI Acc.-No: NC_016810.1) and plasmid (NCBI Acc.-No: NC_017718.1, NC_017719.1, NC_017720.1) reference sequences using the RNA-seq pipeline READemption version 0.3.5 (Förstner *et al*, 2014) and segemehl version 0.2.0 (Hoffmann *et al*, 2014) with an accuracy cutoff of 80%. From the results, only reads mapping uniquely to one genomic position were considered for all subsequent analysis. Pearson correlations between all libraries were calculated on nucleotide read coverage (Appendix Fig S5).

Coverage plots representing the numbers of mapped reads per nt were generated for each replicon and strand to facilitate data visualization in a genome browser. Each resulting cDNA coverage graph was normalized using the DESeq2 (Love *et al*, 2014) size factors calculated during peak calling.

For all analyses related to annotated genomic features such as CDSs, tRNAs, and rRNAs, gene annotations from NCBI were used. We defined *ad hoc* transcriptional units (TUs) based on NCBI CDS annotations, transcription start site (TSS) annotations from Kröger *et al* (2013) and Rho-independent terminator predictions by RNIE (Gardner *et al*, 2011). Briefly, TUs were defined as starting on annotated primary TSSes and ending either with a predicted Rho-independent terminator or in the presence of an intergenic gap greater than 500 nt on the coding strand. In the absence of an upstream TSS, an arbitrary 100 nt 5′UTR was added upstream of the first CDS in the TU, and similarly in the absence of a terminator, an arbitrary 100 nt 3′UTR was added. In the event of a predicted primary TSS within an intergenic gap of less than 500 nt on the coding strand, the TU was ended 100 nt downstream of the preceding CDS, or at the end of the preceding CDS if the predicted primary TSS was less than 100 nt downstream. We defined 5′UTRs as the

Published online: April 4, 2016

regions from the start of each predicted TU to the position upstream of the first CDS in the TU and 3′UTRs as the regions from one nt downstream of the last CDS in the TU to the end of the TU. sRNA annotations are based on Perkins *et al* (2009), Chinni *et al* (2010), Kröger *et al* (2013), and KU Förstner and J Vogel (unpublished data).

Peak calling

Peak calling was performed as a two-step process. In the first step, we defined peak regions using the blockbuster algorithm for defining discrete blocks of overlapping reads (Langenberger *et al*, 2009) across all crosslinked libraries for each RNA-binding protein investigated. Mapped and collapsed reads were filtered to only contain properly paired reads. The resulting BAM files were converted to BED format using BEDTools (v2.17.0) (Quinlan & Hall, 2010). These BED files were concatenated for all crosslinked libraries. Subsequently, each read pair in the concatenated BED file was merged into a single unit representing the sequenced RNA fragment. Only fragments ≤ 25 nt and ≥ 12 nt were retained for further analysis. The resulting BED file was reformatted to satisfy the blockbuster input specifications. Blockbuster uses a greedy approach based on a Gaussian smoothing of read profiles to identify clusters of overlapping read blocks. For this procedure, we required blocks to contain at least 10 reads (i.e., the minBlock-Height option was set to 10) and clusters had to be separated by at least one base (i.e., the distance parameter was set to 1). This procedure resulted in a large set of clusters consisting of overlapping blocks of reads. We then iteratively decomposed each cluster of overlapping blocks into peaks, taking into consideration the local frequency of read counts within the cluster. We first selected the block with the highest read count from the cluster under consideration. All blocks that overlapped with this block were removed from the cluster, and a peak was defined using these overlapping blocks. This procedure, of selecting the next largest block, was repeated in the reduced cluster until no more blocks were left that contained greater than 1% of the total cluster read count (see Appendix Supplementary Methods for a formalized description of this procedure).

In the second step of our peak calling analysis, we applied DESeq2 (v1.2.10) (Love *et al*, 2014) to test each peak for a reproducible relative read count enrichment in triplicate crosslinked libraries compared to non-crosslinked controls. Reads per peak were counted using HTSeq-count (v 0.6.1p1) (Anders *et al*, 2015) for all libraries with the mode option set to “union”, the order option set to “name” and the stranded option set to “yes”. DESeq2 was then run with default options in R. We considered peaks genuine if they had a normalized average expression of ≥ 10 in the crosslinked libraries and a statistically significant enrichment in crosslinked libraries compared to non-crosslinked controls, defined as a false discovery rate (FDR) corrected *P*-value of 0.1 or less.

CopraRNA–Hfq peaks overlap

CopraRNA (Wright *et al*, 2013, 2014) target predictions were performed for all sRNAs from the benchmark dataset of (Wright *et al*, 2013) that had an associated Hfq peak in our data (that is, all except RyhB). Two hundred nucleotides upstream and 100 nucleotides downstream of annotated start codons were specified as potential target regions. The top 20 CopraRNA predictions for each sRNA candidate were subsequently intersected with mRNA candidates that show an Hfq peak in our data. To test for enrichment of known targets in the intersected lists, the number of known targets in the unfiltered top 20 CopraRNA predictions and the number of known targets in the lists resulting from the intersection were compared. The benchmark dataset (Wright *et al*, 2013) was considered as a reference for verified targets and was extended with the interactions between Spot42-glpF (Beisel *et al*, 2012), OxyS-cspC (Tjaden *et al*, 2006), and RybB-STM1530 (Wright, 2012). The unfiltered list of top 20 predictions for 17 individual target predictions contains 51 verified targets in a total list of length 340. The filtered list has a length of 48 and contains 19 verified targets. The interaction between Spot42–mglB discovered in this study was not used for enrichment analysis. A one-sided Fisher’s exact test was employed to test for enrichment of known targets in the filtered list relative to the unfiltered list. The test was performed in R statistics using the Fisher’s exact test function with the “alternative” parameter set to “greater”. For this, we considered that 19 candidates are Hfq bound and verified, 29 candidates are Hfq bound and not verified, 32 candidates are not Hfq bound and verified and 260 candidates are not Hfq bound and not verified. Based on these numbers, the test matrix is given as matrix(c(19,32,29,260), nrow = 2, ncol = 2) in R notation. For the sake of simplicity, we considered targets verified in *E. coli* also to be targets in *Salmonella*. Even though this may not hold true for every single target, this is unlikely to change the principle findings of this analysis.

Analysis of crosslink-specific mutations

For the detection of crosslinking-induced mutation sites from the CLIP-seq data, only uniquely mapped, paired-end reads were considered and used for mutation calling using samtools (v 0.1.19). To reduce bias caused by sequencing errors, we required the mutated sites to be present in both paired reads. A python script adapted from the PIPE-CLIP package (Chen *et al*, 2014) was applied to identify sites significantly enriched in mutations in each library. The number of mutations at each position was modeled as the result of a Bernoulli process with p equal to the observed mutation rate across all positions. Positions were counted as significantly enriched in mutations if the probability of a mutation count greater than or equal to that observed at the position was less than 0.01 under the implied binomial distribution. The final requirement for a site to be considered enriched for crosslinking-induced mutations was that it had to be present in at least two of the libraries from the crosslinked samples and absent in all of the libraries from non-crosslinked samples.

Global analysis of binding regions

The peak density was calculated by counting the number of peaks along the specified annotation features, which included start codons in single-cistron mRNAs and in the first cistron in multigene operons, stop codons in single-cistron mRNAs and in the last cistron in operons, sRNAs, and predicted Rho-independent terminators. These features were retrieved from the extended *Salmonella* Typhimurium SL1344 annotation described above.

Analysis of sequence and structure motifs

The sequences of peaks or sequences 10 nucleotides upstream and downstream of crosslinking mutation sites were used for sequence motif identification using MEME (Bailey *et al*, 2015) with one base shift allowed while the remaining parameters were set at default values. To verify the specificity of the peak motifs found in Hfq peaks from 3′UTRs or sRNAs, the following analysis was performed: for each annotation feature with an Hfq peak, a sequence of the same length as the Hfq peak mapping to that feature but randomly positioned within the feature was extracted. This procedure was repeated ten times. The resulting sequences were used as input for MEME.

To search for the presence of a structural motif, CMfinder 0.2.1 (Yao *et al*, 2006) was run on sequences from peak regions extended by additional 10 nt upstream and downstream, using default parameters except for allowing a minimum single stem loop candidate length of 20 nt. The top-ranked motif incorporated 396 sequences while the motif detected most frequently was found in 416 of the 467 sequences. Both motifs were visualized using R2R (Weinberg & Breaker, 2011) and are depicted in Fig 6F.

Analysis of Hfq peaks in known sRNA–mRNA pairs

Distributions of Hfq peaks in sRNAs and mRNAs with validated basepair interaction sites (Wright *et al*, 2013) were calculated and visualized as a heat map using Excel. The interactions used were restricted to those mRNAs where an Hfq peak was detected within 100 nt on either side of a validated sRNA interaction site.

Pathway analysis

Pathway information was retrieved from the KEGG database (Kanehisa & Goto, 2000), the *Salmonella* SL1344 genome annotation (Kröger *et al*, 2012), and a selection of regulons curated from literature sources. Pathway enrichment analysis was performed using Fisher's exact test, and *P*-values were corrected for multiple testing using the Benjamini–Hochberg method.

Western blot

To analyze immunoprecipitated material in the CLIP experiments, one-tenth of the magnetic beads from each sample was resuspended in 10 μl protein loading buffer and heated 4 min at 95°C. The magnetic beads were collected on a magnetic separator, and the supernatant was loaded and separated on a 15% SDS–polyacrylamide gel followed by transfer of proteins to a nitrocellulose membrane. To detect FLAG-tagged proteins, the membrane was blocked in TBS-T with 5% milk powder, washed in TBS-T for 10 min, incubated for 1 h with anti-FLAG antibody (Sigma) diluted 1:1,000 dilution in TBS-T with 3% BSA, washed in TBS-T for 10 min, incubated for 1 h with anti-mouse-HRP antibody (ThermoScientific) diluted 1:10,000 dilution in TBS-T with 3% BSA, and finally washed in TBS-T two times for 10 min before adding the ECL substrate and taking captions with a CCD camera (ImageQuant, GE Healthcare).

To analyze the expression of GFP fusion proteins, bacterial cultures were harvested at an OD_{600} of 1.0, and cell pellets were boiled in protein loading buffer and separated on 12% SDS–polyacrylamide gels. Proteins were transferred to PVDF membranes and GFP signals were detected as described above but using an anti-GFP antibody (Roche) followed by HRP-coupled anti-mouse antibody (ThermoScientific).

qRT–PCR

Total RNA was extracted using hot phenol, and contaminating DNA was removed by DNase I treatment. qRT–PCRs were carried out using the RNA-to-Ct 1-step kit (ThermoFisher) with 50 ng of RNA per reaction. Relative gene expression was calculated using the $\Delta\Delta C_t$ method (Livak & Schmittgen, 2001) by normalization to the *rfaH* mRNA.

Expanded View for this article is available online.

Acknowledgements

We thank Stan Gorski, Tony Romeo, Gerhart Wagner, and Jay Hinton for comments on the manuscript and Victoria McParland for excellent technical assistance. We also thank Yanjie Chao, Kai Papenfort, Verena Pfeiffer, Dimitri Podkaminski, and Franziska Seifert for providing bacterial strains and plasmids. The Vogel group is supported by funds from DFG and the Bavarian BioSysNet program. Both the Vogel and Backofen group received funds from a joint BMBF eBio:RNASys grant. Erik Holmqvist acknowledges support by an EMBO long-term fellowship and by the Wenner-Gren Foundations. Lars Barquist is supported by a Research Fellowship from the Alexander von Humboldt Foundation.

Author contributions

JV and EH designed the experiments. EH performed the experiments. PRW, TB, and RB implemented the peak calling strategy. LL implemented the calling of crosslink mutations and performed the motif analysis together with TB. PRW, LL, TB, LB, and EH carried out additional bioinformatical analyses. RR carried out sequencing and data analysis. EH and JV wrote the manuscript, which was discussed, modified, and improved by all authors. JV and RB supervised the project.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Altier C, Suyemoto M, Lawhon SD (2000) Regulation of Salmonella enterica serovar typhimurium invasion genes by csrA. *Infect Immun* 68: 6790 – 6797

Anders S, Pyl PT, Huber W (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166 – 169

Andrade JM, Pobre V, Matos AM, Arraiano CM (2012) The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. *RNA* 18: 844 – 855

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

Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T (2012) Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA* 3: 159 – 177

Ascano M, Gerstberger S, Tuschl T (2013) Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. *Curr Opin Genet Dev* 23: 20 – 28

Babitzke P, Baker CS, Romeo T (2009) Regulation of translation initiation by RNA binding proteins. *Annu Rev Microbiol* 63: 27 – 44

Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME Suite. *Nucleic Acids Res* 43: W39 – 49

Baker CS, Morozov I, Suzuki K, Romeo T, Babitzke P (2002) CsrA regulates glycogen biosynthesis by preventing translation of glgC in *Escherichia coli*. *Mol Microbiol* 44: 1599 – 1610

Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* 46: 674 – 690

Bandyra KJ, Said N, Pfeiffer V, Gorna MW, Vogel J, Luisi BF (2012) The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Mol Cell* 47: 943 – 953

Barquist L, Vogel J (2015) Accelerating discovery and functional analysis of small RNAs with new technologies. *Annu Rev Genet* 49: 367 – 394

Beisel CL, Storz G (2011) The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. *Mol Cell* 41: 286 – 297

Beisel CL, Updegrove TB, Janson BJ, Storz G (2012) Multiple factors dictate target selection by Hfq-binding small RNAs. *EMBO J* 31: 1961 – 1974

Bilusic I, Popitsch N, Rescheneder P, Schroeder R, Lybecker M (2014) Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. *RNA Biol* 11: 641 – 654

Brencic A, Lory S (2009) Determination of the regulon and identification of novel mRNA targets of Pseudomonas aeruginosa RsmA. *Mol Microbiol* 72: 612 – 632

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149: 1393 – 1406

Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J (2012) An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* 31: 4005 – 4019

Chao Y, Vogel J (2016) A 3′ UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. *Mol Cell* 61: 352 – 363

Chen B, Yun J, Kim MS, Mendell JT, Xie Y (2014) PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. *Genome Biol* 15: R18

Chinni SV, Raabe CA, Zakaria R, Randau G, Hoe CH, Zemann A, Brosius J, Tang TH, Rozhdestvensky TS (2010) Experimental identification and characterization of 97 novel npcRNA candidates in Salmonella enterica serovar Typhi. *Nucleic Acids Res* 38: 5893 – 5908

Coornaert A, Chiaruttini C, Springer M, Guillier M (2013) Post-transcriptional control of the *Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. *PLoS Genet* 9: e1003156

Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC, Vogel J (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* 84: 428 – 445

Darnell RB (2010) HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA* 1: 266 – 286

De Lay N, Gottesman S (2012) A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. *Mol Microbiol* 86: 524 – 538

Desnoyers G, Masse E (2012) Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes Dev* 26: 726 – 739

Dimastrogiovanni D, Fröhlich KS, Bandyra KJ, Bruce HA, Hohensee S, Vogel J, Luisi BF (2014) Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *ELife* 3: e05375

Dubey AK, Baker CS, Romeo T, Babitzke P (2005) RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* 11: 1579 – 1587

Duss O, Michel E, Diarra Dit Konte N, Schubert M, Allain FH (2014a) Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. *Nucleic Acids Res* 42: 5332 – 5346

Duss O, Michel E, Yulikov M, Schubert M, Jeschke G, Allain FH (2014b) Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* 509: 588 – 592

Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, Camacho MI, Fields JA, Thompson SA, Georgellis D, Cashel M, Babitzke P, Romeo T (2011) Circuitry linking the Csr and stringent response global regulatory systems. *Mol Microbiol* 80: 1561 – 1580

Fender A, Elf J, Hampel K, Zimmermann B, Wagner EG (2010) RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev* 24: 2621 – 2626

Feng L, Rutherford ST, Papenfort K, Bagert JD, van Kessel JC, Tirrell DA, Wingreen NS, Bassler BL (2015) A qrr noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell* 160: 228 – 240

Figueira R, Watson KG, Holden DW, Helaine S (2013) Identification of salmonella pathogenicity island-2 type III secretion system effectors involved in intramacrophage replication of S. enterica serovar typhimurium: implications for rational vaccine design. *MBio* 4: e00065

Figueroa-Bossi N, Lemire S, Maloriol D, Balbontin R, Casadesus J, Bossi L (2006) Loss of Hfq activates the sigmaE-dependent envelope stress response in Salmonella enterica. *Mol Microbiol* 62: 838 – 852

Figueroa-Bossi N, Valentini M, Malleret L, Fiorini F, Bossi L (2009) Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. *Genes Dev* 23: 2004 – 2015

Figueroa-Bossi N, Schwartz A, Guillemardet B, D'Heygere F, Bossi L, Boudvillain M (2014) RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev* 28: 1239 – 1251

Förstner KU, Vogel J, Sharma CM (2014) READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics* 30: 3421 – 3423

Friedersdorf MB, Keene JD (2014) Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome Biol* 15: R2

Gardner PP, Barquist L, Bateman A, Nawrocki EP, Weinberg Z (2011) RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res* 39: 5845 – 5852

Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA (2011) Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc Natl Acad Sci USA* 108: 12875 – 12880

Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, Storz G (2014) MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev* 28: 1620 – 1634

Hajnsdorf E, Regnier P (2000) Host factor Hfq of *Escherichia coli* stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proc Natl Acad Sci USA* 97: 1501 – 1505

Hébrard M, Kröger C, Srikumar S, Colgan A, Händler K, Hinton JC (2012) sRNAs and the virulence of Salmonella enterica serovar Typhimurium. *RNA Biol* 9: 437 – 445

Hensel M (2004) Evolution of pathogenicity islands of Salmonella enterica. *Int J Med Microbiol* 294: 95 – 102

Heroven AK, Bohme K, Dersch P (2012) The Csr/Rsm system of Yersinia and related pathogens: a post-transcriptional strategy for managing virulence. *RNA Biol* 9: 379 – 391

Hoffmann S, Otto C, Doose G, Tanzer A, Langenberger D, Christ S, Kunz M, Holdt LM, Teupser D, Hackermuller J, Stadler PF (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol* 15: R34

Holmqvist E, Reimegård J, Sterk M, Grantcharova N, Römling U, Wagner EGH (2010) Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J* 29: 1840 – 1850

Holmqvist E, Vogel J (2013) A small RNA serving both the Hfq and CsrA regulons. *Genes Dev* 27: 1073 – 1078

Ishikawa H, Otaka H, Maki K, Morita T, Aiba H (2012) The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3′ poly(U) tail. *RNA* 18: 1062 – 1074

Jørgensen MG, Nielsen JS, Boysen A, Franch T, Møller-Jensen J, Valentin-Hansen P (2012) Small regulatory RNAs control the multi-cellular adhesive lifestyle of *Escherichia coli*. *Mol Microbiol* 84: 36 – 50

Jørgensen MG, Thomason MK, Havelund J, Valentin-Hansen P, Storz G (2013) Dual function of the McaS small RNA in controlling biofilm formation. *Genes Dev* 27: 1132 – 1145

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27 – 30

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059 – 3066

Kawamoto H, Koide Y, Morita T, Aiba H (2006) Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol* 61: 1013 – 1022

Kawano M, Reynolds AA, Miranda-Rios J, Storz G (2005) Detection of 5′- and 3′-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res* 33: 1040 – 1050

König J, Zarnack K, Luscombe NM, Ule J (2011) Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet* 13: 77 – 83

Koo JT, Alleyne TM, Schiano CA, Jafari N, Lathem WW (2011) Global discovery of small RNAs in Yersinia pseudotuberculosis identifies Yersinia-specific small, noncoding RNAs required for virulence. *Proc Natl Acad Sci USA* 108: E709 – E717

Kramer K, Sachsenberg T, Beckmann BM, Qamar S, Boon KL, Hentze MW, Kohlbacher O, Urlaub H (2014) Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods* 11: 1064 – 1070

Kröger C, Dillon SC, Cameron AD, Papenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hebrard M, Handler K, Colgan A, Leekitcharoenphon P, Langridge GC, Lohan AJ, Loftus B, Lucchini S, Ussery DW, Dorman CJ, Thomson NR, Vogel J *et al* (2012) The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. *Proc Natl Acad Sci USA* 109: E1277 – 1286

Kröger C, Colgan A, Srikumar S, Handler K, Sivasankaran SK, Hammarlof DL, Canals R, Grissom JE, Conway T, Hokamp K, Hinton JC (2013) An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe* 14: 683 – 695

Lalaouna D, Carrier MC, Semsey S, Brouard JS, Wang J, Wade JT, Masse E (2015) A 3′ external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Mol Cell* 58: 393 – 405

Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF (2009) Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics* 25: 2298 – 2301

Lapouge K, Perozzo R, Iwaszkiewicz J, Bertelli C, Zoete V, Michielin O, Scapozza L, Haas D (2013) RNA pentaloop structures as effective targets of regulators belonging to the RsmA/CsrA protein family. *RNA Biol* 10: 1031 – 1041

Lawhon SD, Frye JG, Suyemoto M, Porwollik S, McClelland M, Altier C (2003) Global regulation by CsrA in Salmonella typhimurium. *Mol Microbiol* 48: 1633 – 1645

Le Derout J, Folichon M, Briani F, Deho G, Regnier P, Hajnsdorf E (2003) Hfq affects the length and the frequency of short oligo(A) tails at the 3′ end of *Escherichia coli* rpsO mRNAs. *Nucleic Acids Res* 31: 4017 – 4023

Lease RA, Woodson SA (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J Mol Biol* 344: 1211 – 1223

Lenz DH, Miller MB, Zhu J, Kulkarni RV, Bassler BL (2005) CsrA and three redundant small RNAs regulate quorum sensing in Vibrio cholerae. *Mol Microbiol* 58: 1186 – 1202

Link TM, Valentin-Hansen P, Brennan RG (2009) Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci USA* 106: 19292 – 19297

Liu MY, Yang H, Romeo T (1995) The product of the pleiotropic *Escherichia coli* gene csrA modulates glycogen biosynthesis via effects on mRNA stability. *J Bacteriol* 177: 2663 – 2672

Liu MY, Gui G, Wei B, Preston JF III, Oakford L, Yuksel U, Giedroc DP, Romeo T (1997) The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem* 272: 17502 – 17510

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402 – 408

Lopez-Garrido J, Puerta-Fernandez E, Casadesus J (2014) A eukaryotic-like 3′ untranslated region in Salmonella enterica hilD mRNA. *Nucleic Acids Res* 42: 5894 – 5906

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550

Majdalani N, Vanderpool CK, Gottesman S (2005) Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol* 40: 93 – 113

Mark Glover JN, Chaulk SG, Edwards RA, Arthur D, Lu J, Frost LS (2015) The FinO family of bacterial RNA chaperones. *Plasmid* 78: 79 – 87

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17: 10 – 12

Martinez LC, Yakhnin H, Camacho MI, Georgellis D, Babitzke P, Puente JL, Bustamante VH (2011) Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the Salmonella SPI-1 and SPI-2 virulence regulons through HilD. *Mol Microbiol* 80: 1637 – 1656

Mercante J, Suzuki K, Cheng X, Babitzke P, Romeo T (2006) Comprehensive alanine-scanning mutagenesis of *Escherichia coli* CsrA defines two subdomains of critical functional importance. *J Biol Chem* 281: 31832 – 31842

Mika F, Busse S, Possling A, Berkholz J, Tschowri N, Sommerfeldt N, Pruteanu M, Hengge R (2012) Targeting of csgD by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in *Escherichia coli*. *Mol Microbiol* 84: 51 – 65

Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, Feig AL (2004) *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol* 11: 1206 – 1214

Miyakoshi M, Chao Y, Vogel J (2015a) Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J* 34: 1478 – 1492

Miyakoshi M, Chao Y, Vogel J (2015b) Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Curr Opin Microbiol* 24: 132 – 139

Møller T, Franch T, Hojrup P, Keene DR, Bachinger HP, Brennan RG, Valentin-Hansen P (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell* 9: 23 – 30

Morita T, Aiba H (2011) RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. *Genes Dev* 25: 294 – 298

Otaka H, Ishikawa H, Morita T, Aiba H (2011) PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc Natl Acad Sci USA* 108: 13059 – 13064

Panja S, Schu DJ, Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. *Nucleic Acids Res* 41: 7536 – 7546

Papenfort K, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J (2013) Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* 153: 426 – 437

Papenfort K, Förstner KU, Cong JP, Sharma CM, Bassler BL (2015) Differential RNA-seq of Vibrio cholerae identifies the VqmR small RNA as a regulator of biofilm formation. *Proc Natl Acad Sci USA* 112: E766 – 775

Peng Y, Soper TJ, Woodson SA (2014) Positional effects of AAN motifs in rpoS regulation by sRNAs and Hfq. *J Mol Biol* 426: 275 – 285

Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomson NR, Dougan G (2009) A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus Salmonella typhi. *PLoS Genet* 5: e1000569

Phadtare S, Alsina J, Inouye M (1999) Cold-shock response and cold-shock proteins. *Curr Opin Microbiol* 2: 175 – 180

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841 – 842

Rice JB, Balasubramanian D, Vanderpool CK (2012) Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA. *Proc Natl Acad Sci USA* 109: E2691 – 2698

Romeo T, Gong M, Liu MY, Brun-Zinkernagel AM (1993) Identification and molecular characterization of csrA, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J Bacteriol* 175: 4744 – 4755

Romeo T, Vakulskas CA, Babitzke P (2013) Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environ Microbiol* 15: 313 – 324

Salim NN, Feig AL (2010) An upstream Hfq binding site in the fhlA mRNA leader region facilitates the OxyS-fhlA interaction. *PLoS ONE* 5: e13028

Salim NN, Faner MA, Philip JA, Feig AL (2012) Requirement of upstream Hfq-binding (ARN)x elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY. *Nucleic Acids Res* 40: 8021 – 8032

Sauer E, Weichenrieder O (2011) Structural basis for RNA 3′-end recognition by Hfq. *Proc Natl Acad Sci USA* 108: 13065 – 13070

Sauer E, Schmidt S, Weichenrieder O (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc Natl Acad Sci USA* 109: 9396 – 9401

Schu DJ, Zhang A, Gottesman S, Storz G (2015) Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J* 34: 2557 – 2573

Schumacher MA, Pearson RF, Möller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J* 21: 3546 – 3556

Sittka A, Pfeiffer V, Tedin K, Vogel J (2007) The RNA chaperone Hfq is essential for the virulence of Salmonella typhimurium. *Mol Microbiol* 63: 193 – 217

Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* 4: e1000163

Soper TJ, Doxzen K, Woodson SA (2011) Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *RNA* 17: 1544 – 1550

Sterzenbach T, Nguyen KT, Nuccio SP, Winter MG, Vakulskas CA, Clegg S, Romeo T, Bäumler AJ (2013) A novel CsrA titration mechanism regulates fimbrial gene expression in Salmonella typhimurium. *EMBO J* 32: 2872 – 2883

Sugimoto Y, König J, Hussain S, Zupan B, Curk T, Frye M, Ule J (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol* 13: R67

Thomason MK, Fontaine F, De Lay N, Storz G (2012) A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol Microbiol* 84: 17 – 35

Tjaden B, Goodwin SS, Opdyke JA, Guillier M, Fu DX, Gottesman S, Storz G (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res* 34: 2791 – 2802

Tree JJ, Granneman S, McAteer SP, Tollervey D, Gally DL (2014) Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*. *Mol Cell* 55: 199 – 213

Updegrove TB, Wartell RM (2011) The influence of *Escherichia coli* Hfq mutations on RNA binding and sRNA*mRNA duplex formation in rpoS riboregulation. *Biochim Biophys Acta* 1809: 532 – 540

Updegrove TB, Shabalina SA, Storz G (2015) How do base-pairing small RNAs evolve? *FEMS Microbiol Rev* 39: 379 – 391

Uren PJ, Bahrami-Samani E, Burns SC, Qiao M, Karginov FV, Hodges E, Hannon GJ, Sanford JR, Penalva LO, Smith AD (2012) Site identification in high-throughput RNA-protein interaction data. *Bioinformatics* 28: 3013 – 3020

Vakulskas CA, Potts AH, Babitzke P, Ahmer BM, Romeo T (2015) Regulation of bacterial virulence by Csr (Rsm) systems. *Microbiol Mol Biol Rev* 79: 193 – 224

Valverde C, Lindell M, Wagner EG, Haas D (2004) A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of Pseudomonas fluorescens. *J Biol Chem* 279: 25066 – 25074

Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jäger JG, Hüttenhofer A, Wagner EG (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* 31: 6435 – 6443

Vogel J (2009) A rough guide to the non-coding RNA world of Salmonella. *Mol Microbiol* 71: 1 – 11

Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* 9: 578 – 589

Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, Baker CS, Georgellis D, Babitzke P, Romeo T (2003) A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol Microbiol* 48: 657 – 670

Weinberg Z, Breaker RR (2011) R2R–software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* 12: 3

Westermann AJ, Förstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Müller L, Reinhardt R, Stadler PF, Vogel J (2016) Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* 529: 496 – 501

Wilson KS, von Hippel PH (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc Natl Acad Sci USA* 92: 8793 – 8797

Wilusz CJ, Wilusz J (2005) Eukaryotic Lsm proteins: lessons from bacteria. *Nat Struct Mol Biol* 12: 1031 – 1036

Wright PR (2012) hIntaRNA – Comparative prediction of sRNA targets in prokaryotes. Diploma, Albert Ludwig University Freiburg, Freiburg, Germany

Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proc Natl Acad Sci USA* 110: E3487 – E3496

Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, Kleinkauf R, Hess WR, Backofen R (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res* 42: W119 – 123

Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S (2012) FastUniq: a fast *de novo* duplicates removal tool for paired short reads. *PLoS ONE* 7: e52249

Yakhnin AV, Baker CS, Vakulskas CA, Yakhnin H, Berezin I, Romeo T, Babitzke P (2013) CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. *Mol Microbiol* 87: 851 – 866

Yao Z, Weinberg Z, Ruzzo WL (2006) CMfinder–a covariance model based RNA motif finding algorithm. *Bioinformatics* 22: 445 – 452

Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell* 9: 11 – 22

Zhang C, Darnell RB (2011) Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* 29: 607 – 614

Zhang A, Schu DJ, Tjaden BC, Storz G, Gottesman S (2013) Mutations in interaction surfaces differentially impact *E. coli* Hfq association with small RNAs and their mRNA targets. *J Mol Biol* 425: 3678 – 3697

Zheng D, Constantinidou C, Hobman JL, Minchin SD (2004) Identification of the CRP regulon using *in vitro* and *in vivo* transcriptional profiling. *Nucleic Acids Res* 32: 5874 – 5893

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406 – 3415

9 Bibliography

[1] R. E. Franklin and R. G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171(4356):740–1, 1953.

[2] J. D. Watson and F. H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–8, 1953.

[3] F. Sanger, S. Nicklen, and A. R. Coulson. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*, 74(12):5463–7, 1977.

[4] S. Goodwin, J. D. McPherson, and W. R. McCombie. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, 17(6):333–51, 2016.

[5] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlen, and P. Nyren. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem*, 242(1):84–9, 1996.

[6] P. Nyren. The history of pyrosequencing. *Methods Mol Biol*, 373:1–14, 2007.

[7] D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. J. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. D. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, M. Chiara E Catenazzi, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, and D. L. Gustafson, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218):53–9, 2008.

[8] W. Gilbert. The RNA world. *Nature*, 319:618, 1986.

[9] A. Serganov and D. J. Patel. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet*, 8(10):776–90, 2007.

[10] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–3, 1970.

[11] B. Lewin. *Genes 9*. Jones & Bartlett Learning, 2008.

[12] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.

[13] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 5(7):621–8, 2008.

[14] Z. Wang, M. Gerstein, and M. Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10(1):57–63, 2009.

[15] E. Pennisi. Genomics. ENCODE project writes eulogy for junk DNA. *Science*, 337(6099):1159, 1161, 2012.

[16] J. Ortin and F. Parra. Structure and function of RNA replication. *Annu Rev Microbiol*, 60:305–26, 2006.

[17] W.-S. Hu and S. H. Hughes. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med*, 2(10), 2012.

[18] S. B. Cohen, M. E. Graham, G. O. Lovrecz, N. Bache, P. J. Robinson, and R. R. Reddel. Protein composition of catalytically active human telomerase from immortal cells. *Science*, 315(5820):1850–3, 2007.

[19] L. Salmena, L. Poliseno, Y. Tay, L. Kats, and P. P. Pandolfi. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, 146(3):353–8, 2011.

[20] A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, and S. A. Teichmann. The technology and biology of single-cell RNA sequencing. *Mol Cell*, 58(4):610–20, 2015.

[21] A.-E. Saliba, L. Li, A. J. Westermann, S. Appenzeller, D. A. C. Stapels, L. N. Schulte, S. Helaine, and J. Vogel. Single-cell RNA-seq ties macrophage polarization to growth rate of intracellular *Salmonella*. *Nat Microbiol*, 2:16206, 2016.

[22] S. Gottesman and G. Storz. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol*, 3(12):a003798, 2011.

[23] G. Storz, J. Vogel, and K. M. Wassarman. Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell*, 43(6):880–91, 2011.

[24] L. Barquist and J. Vogel. Accelerating Discovery and Functional Analysis of Small RNAs with New Technologies. *Annu Rev Genet*, 49:367–94, 2015.

[25] E. G. H. Wagner and P. Romby. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv Genet*, 90:133–208, 2015.

[26] H. Gram. Ueber die isolierte faerbung der schizomyceten in schnitt- und trockenpraeparaten. *Fortschritte der Medizin*, 2:185–89, 1884.

[27] J. Hinnebusch and K. Tilly. Linear plasmids and chromosomes in bacteria. *Mol Microbiol*, 10(5):917–22, 1993.

[28] F. Pfeifer. Distribution, formation and regulation of gas vesicles. *Nat Rev Microbiol*, 10(10):705–15, 2012.

[29] J. M. Shively. Inclusion bodies of prokaryotes. *Annu Rev Microbiol*, 28(0):167–87, 1974.

[30] K. D. Young. The selective value of bacterial shape. *Microbiol Mol Biol Rev*, 70(3):660–703, 2006.

[31] J. A. Shapiro. Thinking about bacterial populations as multicellular organisms. *Annu Rev Microbiol*, 52:81–104, 1998.

[32] K. Kumar, R. A. Mella-Herrera, and J. W. Golden. Cyanobacterial heterocysts. *Cold Spring Harb Perspect Biol*, 2(4):a000315, 2010.

[33] E. Flores and A. Herrero. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nat Rev Microbiol*, 8(1):39–50, 2010.

[34] J. D. Wang and P. A. Levin. Metabolism, cell growth and the bacterial cell cycle. *Nat Rev Microbiol*, 7(11):822–7, 2009.

[35] E. F. Bi and J. Lutkenhaus. FtsZ ring structure associated with division in *Escherichia coli*. *Nature*, 354(6349):161–4, 1991.

[36] E. Harry, L. Monahan, and L. Thompson. Bacterial cell division: the mechanism and its precison. *Int Rev Cytol*, 253:27–94, 2006.

[37] J. Pommerville. *Alcamo's Fundamentals of Microbiology*. Jones & Bartlett Learning, 2010.

[38] J. Errington. Regulation of endospore formation in *Bacillus subtilis*. *Nat Rev Microbiol*, 1(2):117–26, 2003.

[39] A. Checinska, A. Paszczynski, and M. Burbank. *Bacillus* and other spore-forming genera: variations in responses and mechanisms for survival. *Annu Rev Food Sci Technol*, 6:351–69, 2015.

[40] R. J. Cano and M. K. Borucki. Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber. *Science*, 268(5213):1060–4, 1995.

[41] C. S. Ting, G. Rocap, J. King, and S. W. Chisholm. Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol*, 10(3):134–42, 2002.

[42] A. Reyes-Prieto, A. P. M. Weber, and D. Bhattacharya. The origin and estab-

lishment of the plastid in algae and plants. *Annu Rev Genet*, 41:147–68, 2007.

[43] N. Schuergers, T. Lenn, R. Kampmann, M. V. Meissner, T. Esteves, M. Temerinac-Ott, J. G. Korvink, A. R. Lowe, C. W. Mullineaux, and A. Wilde. Cyanobacteria use micro-optics to sense light direction. *Elife*, 5, 2016.

[44] M. B. Miller and B. L. Bassler. Quorum sensing in bacteria. *Annu Rev Microbiol*, 55:165–99, 2001.

[45] K. Papenfort and B. L. Bassler. Quorum sensing signal-response systems in Gram-negative bacteria. *Nat Rev Microbiol*, 14(9):576–88, 2016.

[46] A. Fleming. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. 1929. *Bull World Health Organ*, 79(8):780–90, 2001.

[47] A.-P. Magiorakos, A. Srinivasan, R. B. Carey, Y. Carmeli, M. E. Falagas, C. G. Giske, S. Harbarth, J. F. Hindler, G. Kahlmeter, B. Olsson-Liljequist, D. L. Paterson, L. B. Rice, J. Stelling, M. J. Struelens, A. Vatopoulos, J. T. Weber, and D. L. Monnet. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect*, 18(3):268–81, 2012.

[48] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819):1709–12, 2007.

[49] S. J. Lange, O. S. Alkhnbashi, D. Rose, S. Will, and R. Backofen. CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res*, 41(17):8034–44, 2013. SJL, OSA and DR contributed equally to this work.

[50] J. A. Fuerst. Intracellular compartmentation in planctomycetes. *Annu Rev Microbiol*, 59:299–328, 2005.

[51] M. Pilhofer, K. Rappl, C. Eckl, A. P. Bauer, W. Ludwig, K.-H. Schleifer, and G. Petroni. Characterization and evolution of cell division and cell wall synthesis genes in the bacterial phyla *Verrucomicrobia*, *Lentisphaerae*, *Chlamydiae*, and *Planctomycetes* and phylogenetic comparison with rRNA genes. *J Bacteriol*, 190(9):3192–202, 2008.

[52] C. Frank, D. Werber, J. P. Cramer, M. Askar, M. Faber, M. an der Heiden, H. Bernard, A. Fruth, R. Prager, A. Spode, M. Wadl, A. Zoufaly, S. Jordan, M. J. Kemper, P. Follin, L. Muller, L. A. King, B. Rosner, U. Buchholz, K. Stark, and G. Krause. Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4

outbreak in Germany. *N Engl J Med*, 365(19):1771–80, 2011.

[53] L. A. King, F. Nogareda, F.-X. Weill, P. Mariani-Kurkdjian, E. Loukiadis, G. Gault, N. Jourdan-DaSilva, E. Bingen, M. Mace, D. Thevenot, N. Ong, C. Castor, H. Noel, D. Van Cauteren, M. Charron, V. Vaillant, B. Aldabe, V. Goulet, G. Delmas, E. Couturier, Y. Le Strat, C. Combe, Y. Delmas, F. Terrier, B. Vendrely, P. Rolland, and H. de Valk. Outbreak of Shiga toxin-producing *Escherichia coli* O104:H4 associated with organic fenugreek sprouts, France, June 2011. *Clin Infect Dis*, 54(11):1588–94, 2012.

[54] J. M. Rangel, P. H. Sparling, C. Crowe, P. M. Griffin, and D. L. Swerdlow. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982-2002. *Emerg Infect Dis*, 11(4):603–9, 2005.

[55] J. B. Kaper, J. P. Nataro, and H. L. Mobley. Pathogenic *Escherichia coli*. *Nat Rev Microbiol*, 2(2):123–40, 2004.

[56] M. A. Croxen and B. B. Finlay. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol*, 8(1):26–38, 2010.

[57] G. Bertani. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J Bacteriol*, 186(3):595–600, 2004.

[58] R. Bentley and R. Meganathan. Biosynthesis of vitamin K (menaquinone) in bacteria. *Microbiol Rev*, 46(3):241–80, 1982.

[59] S. Hudault, J. Guignot, and A. L. Servin. *Escherichia coli* strains colonising the gastrointestinal tract protect germfree mice against *Salmonella* typhimurium infection. *Gut*, 49(1):47–55, 2001.

[60] F. R. Blattner, G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331):1453–62, 1997.

[61] I. M. Keseler, A. Mackie, M. Peralta-Gil, A. Santos-Zavaleta, S. Gama-Castro, C. Bonavides-Martinez, C. Fulcher, A. M. Huerta, A. Kothari, M. Krummenacker, M. Latendresse, L. Muniz-Rascado, Q. Ong, S. Paley, I. Schroder, A. G. Shearer, P. Subhraveti, M. Travers, D. Weerasinghe, V. Weiss, J. Collado-Vides, R. P. Gunsalus, I. Paulsen, and P. D. Karp. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res*, 41(Database issue):D605–12, 2013.

[62] J. Lederberg and E. L. Tatum. Gene recombination in *Escherichia coli*. *Nature*, 158(4016):558, 1946.

[63] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, 3:318–56, 1961.

[64] S. N. Cohen, A. C. Chang, H. W. Boyer,

and R. B. Helling. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci USA*, 70(11):3240–4, 1973.

[65] A. C. Chang and S. N. Cohen. Genome construction between bacterial species in vitro: replication and expression of *Staphylococcus* plasmid genes in *Escherichia coli*. *Proc Natl Acad Sci USA*, 71(4):1030–4, 1974.

[66] J. F. Morrow, S. N. Cohen, A. C. Chang, H. W. Boyer, H. M. Goodman, and R. B. Helling. Replication and transcription of eukaryotic DNA in *Escherichia coli*. *Proc Natl Acad Sci USA*, 71(5):1743–7, 1974.

[67] R. Crea, A. Kraszewski, T. Hirose, and K. Itakura. Chemical synthesis of genes for human insulin. *Proc Natl Acad Sci USA*, 75(12):5765–9, 1978.

[68] D. V. Goeddel, D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A. D. Riggs. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc Natl Acad Sci USA*, 76(1):106–10, 1979.

[69] J. Ihssen, M. Kowarik, S. Dilettoso, C. Tanner, M. Wacker, and L. Thony-Meyer. Production of glycoprotein vaccines in *Escherichia coli*. *Microb Cell Fact*, 9:61, 2010.

[70] T. Liu and C. Khosla. Genetic engineering of *Escherichia coli* for biofuel production. *Annu Rev Genet*, 44:53–69, 2010.

[71] J. H. Urban and J. Vogel. Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res*, 35(3):1018–37, 2007.

[72] J. H. Urban and J. Vogel. A Green Fluorescent Protein (GFP)-Based Plasmid System to Study Post-Transcriptional Control of Gene Expression In Vivo. *Methods Mol Biol*, 540:301–19, 2009.

[73] L. Fantappie, F. Oriente, A. Muzzi, D. Serruto, V. Scarlato, and I. Delany. A novel Hfq-dependent sRNA that is under FNR control and is synthesized in oxygen limitation in *Neisseria meningitidis*. *Mol Microbiol*, 80(2):507–23, 2011.

[74] A. S. Richter, C. Schleberger, R. Backofen, and C. Steglich. Seed-based IntaRNA prediction combined with GFP-reporter system identifies mRNA targets of the small RNA Yfr1. *Bioinformatics*, 26(1):1–5, 2010.

[75] J. Georg, D. Dienst, N. Schurgers, T. Wallner, D. Kopp, D. Stazic, E. Kuchmina, S. Klahn, H. Lokstein, W. R. Hess, and A. Wilde. The Small Regulatory RNA SyR1/PsrR1 Controls Photosynthetic Functions in Cyanobacteria. *Plant Cell*, 26(9):3661–79, 2014.

[76] S. Klahn, C. Schaal, J. Georg, D. Baumgartner, G. Knippen, M. Hagemann, A. M.

Muro-Pastor, and W. R. Hess. The sRNA NsiR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7. *Proc Natl Acad Sci USA*, 112(45):E6243–52, 2015.

[77] P. R. Wright, A. S. Richter, K. Papenfort, M. Mann, J. Vogel, W. R. Hess, R. Backofen, and J. Georg. Comparative genomics boosts target prediction for bacterial small RNAs. *Proc Natl Acad Sci USA*, 110(37):E3487–96, 2013.

[78] B. Coburn, G. A. Grassl, and B. B. Finlay. *Salmonella*, the host and disease: a brief review. *Immunol Cell Biol*, 85(2):112–8, 2007.

[79] C. Kröger, S. C. Dillon, A. D. S. Cameron, K. Papenfort, S. K. Sivasankaran, K. Hokamp, Y. Chao, A. Sittka, M. Hébrard, K. Händler, A. Colgan, P. Leekitcharoenphon, G. C. Langridge, A. J. Lohan, B. Loftus, S. Lucchini, D. W. Ussery, C. J. Dorman, N. R. Thomson, J. Vogel, and J. C. D. Hinton. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci USA*, 109(20):E1277–86, 2012.

[80] J. A. Crump, S. P. Luby, and E. D. Mintz. The global burden of typhoid fever. *Bull World Health Organ*, 82(5):346–53, 2004.

[81] C. M. Sharma, F. Darfeuille, T. H. Plantinga, and J. Vogel. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev*, 21(21):2804–17, 2007.

[82] M. Bouvier, C. M. Sharma, F. Mika, K. H. Nierhaus, and J. Vogel. Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol Cell*, 32(6):827–37, 2008.

[83] K. Papenfort, N. Said, T. Welsink, S. Lucchini, J. C. D. Hinton, and J. Vogel. Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol Microbiol*, 74(1):139–58, 2009.

[84] P. Mandin and S. Gottesman. Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO J*, 29(18):3094–107, 2010.

[85] K. Papenfort, M. Bouvier, F. Mika, C. M. Sharma, and J. Vogel. Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc Natl Acad Sci USA*, 107(47):20435–40, 2010.

[86] R. Balbontín, F. Fiorini, N. Figueroa-Bossi, J. Casadesús, and L. Bossi. Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in *Salmonella enterica*. *Mol Microbiol*, 78(2):380–94, 2010.

[87] C. M. Sharma, K. Papenfort, S. R. Pernitzsch, H.-J. Mollenkopf, J. C. D. Hin-

ton, and J. Vogel. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol*, 81(5):1144–65, 2011.

[88] P. R. Wright. hIntaRNA – Comparative prediction of sRNA targets in prokaryotes. Diplomarbeit, Albert Ludwigs University Freiburg, March 2012.

[89] M. McClelland, K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, 413(6858):852–6, 2001.

[90] E. A. Groisman and H. Ochman. Cognate gene clusters govern invasion of host epithelial cells by *Salmonella typhimurium* and *Shigella flexneri*. *EMBO J*, 12(10):3779–87, 1993.

[91] E. A. Groisman and H. Ochman. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell*, 87(5):791–4, 1996.

[92] D. M. Mills, V. Bajaj, and C. A. Lee. A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. *Mol Microbiol*, 15(4):749–59, 1995.

[93] J. E. Galan. Molecular genetic bases of *Salmonella* entry into host cells. *Mol Microbiol*, 20(2):263–71, 1996.

[94] J. E. Shea, M. Hensel, C. Gleeson, and D. W. Holden. Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*. *Proc Natl Acad Sci USA*, 93(6):2593–7, 1996.

[95] H. Ochman, F. C. Soncini, F. Solomon, and E. A. Groisman. Identification of a pathogenicity island required for *Salmonella* survival in host cells. *Proc Natl Acad Sci USA*, 93(15):7800–4, 1996.

[96] A. J. Westermann, K. U. Forstner, F. Amman, L. Barquist, Y. Chao, L. N. Schulte, L. Muller, R. Reinhardt, P. F. Stadler, and J. Vogel. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature*, 529(7587):496–501, 2016.

[97] C. Ahrens. *Essentials of Meteorology: An Invitation to the Atmosphere*. Cengage Learning, 2011.

[98] W. Hopkins and N. Hüner. *Introduction to plant physiology*. Wiley, 2008.

[99] S. R. Long. Rhizobium-legume nodulation: life together in the underground. *Cell*, 56(2):203–14, 1989.

[100] R. Op den Camp, A. Streng, S. De Mita, Q. Cao, E. Polone, W. Liu, J. S. S. Ammiraju, D. Kudrna, R. Wing, A. Untergasser,

T. Bisseling, and R. Geurts. LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science*, 331(6019):909–12, 2011.

[101] Wall. The Actinorhizal Symbiosis. *J Plant Growth Regul*, 19(2):167–182, 2000.

[102] D. G. Bullock. Crop rotation. *Critical Reviews in Plant Sciences*, 11(4):309–326, 1992.

[103] D. Capela, F. Barloy-Hubler, J. Gouzy, G. Bothe, F. Ampe, J. Batut, P. Boistard, A. Becker, M. Boutry, E. Cadieu, S. Dreano, S. Gloux, T. Godrie, A. Goffeau, D. Kahn, E. Kiss, V. Lelaure, D. Masuy, T. Pohl, D. Portetelle, A. Puhler, B. Purnelle, U. Ramsperger, C. Renard, P. Thebault, M. Vandenbol, S. Weidner, and F. Galibert. Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc Natl Acad Sci USA*, 98(17):9877–82, 2001.

[104] F. Galibert, T. M. Finan, S. R. Long, A. Puhler, P. Abola, F. Ampe, F. Barloy-Hubler, M. J. Barnett, A. Becker, P. Boistard, G. Bothe, M. Boutry, L. Bowser, J. Buhrmester, E. Cadieu, D. Capela, P. Chain, A. Cowie, R. W. Davis, S. Dreano, N. A. Federspiel, R. F. Fisher, S. Gloux, T. Godrie, A. Goffeau, B. Golding, J. Gouzy, M. Gurjal, I. Hernandez-Lucas, A. Hong, L. Huizar, R. W. Hyman, T. Jones, D. Kahn, M. L. Kahn, S. Kalman, D. H. Keating, E. Kiss, C. Komp, V. Lelaure, D. Masuy, C. Palm, M. C. Peck, T. M. Pohl, D. Portetelle, B. Purnelle, U. Ramsperger, R. Surzycki, P. Thebault, M. Vandenbol, F. J. Vorholter, S. Weidner, D. H. Wells, K. Wong, K. C. Yeh, and J. Batut. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science*, 293(5530):668–72, 2001.

[105] P. Roche, F. Maillet, C. Plazanet, F. Debelle, M. Ferro, G. Truchet, J. C. Prome, and J. Denarie. The common nodABC genes of *Rhizobium meliloti* are host-range determinants. *Proc Natl Acad Sci USA*, 93(26):15305–10, 1996.

[106] K. M. Jones, H. Kobayashi, B. W. Davies, M. E. Taga, and G. C. Walker. How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nat Rev Microbiol*, 5(8):619–33, 2007.

[107] B. A. Webb, S. Hildreth, R. F. Helm, and B. E. Scharf. *Sinorhizobium meliloti* chemoreceptor McpU mediates chemotaxis toward host plant exudates through direct proline sensing. *Appl Environ Microbiol*, 80(11):3404–15, 2014.

[108] X. Perret, C. Staehelin, and W. J. Broughton. Molecular basis of symbiotic promiscuity. *Microbiol Mol Biol Rev*, 64(1):180–201, 2000.

[109] G. E. D. Oldroyd and J. A. Downie. Calcium, kinases and nodulation signalling in

legumes. *Nat Rev Mol Cell Biol*, 5(7):566–76, 2004.

[110] J. J. Esseling, F. G. P. Lhuissier, and A. M. C. Emons. Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application. *Plant Physiol*, 132(4):1982–8, 2003.

[111] D. J. Gage. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol Mol Biol Rev*, 68(2):280–300, 2004.

[112] H. M. Fischer. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol Rev*, 58(3):352–86, 1994.

[113] R. M. Mitra and S. R. Long. Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the *Medicago truncatula/Sinorhizobium meliloti* symbiosis. *Plant Physiol*, 134(2):595–604, 2004.

[114] C. G. Starker, A. L. Parra-Colmenares, L. Smith, R. M. Mitra, and S. R. Long. Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. *Plant Physiol*, 140(2):671–80, 2006.

[115] C. I. Pislariu and R. Dickstein. An IRE-like AGC kinase gene, MtIRE, has unique expression in the invasion zone of developing root nodules in *Medicago truncatula*. *Plant Physiol*, 144(2):682–94, 2007.

[116] R. Dixon and D. Kahn. Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol*, 2(8):621–31, 2004.

[117] T. Ott, J. T. van Dongen, C. Gunther, L. Krusell, G. Desbrosses, H. Vigeolas, V. Bock, T. Czechowski, P. Geigenberger, and M. K. Udvardi. Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Curr Biol*, 15(6):531–5, 2005.

[118] D. B. Lobell, W. Schlenker, and J. Costa-Roberts. Climate trends and global crop production since 1980. *Science*, 333(6042):616–20, 2011.

[119] M. Robledo, B. Frage, P. R. Wright, and A. Becker. A stress-induced small RNA modulates alpha-rhizobial cell cycle progression. *PLoS Genet*, 11(4):e1005153, 2015.

[120] J.-P. Schluter, J. Reinkensmeier, S. Daschkey, E. Evguenieva-Hackenberg, S. Janssen, S. Janicke, J. D. Becker, R. Giegerich, and A. Becker. A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genomics*, 11:245, 2010.

[121] D. W. Wood, J. C. Setubal, R. Kaul, D. E. Monks, J. P. Kitajima, V. K. Okura, Y. Zhou, L. Chen, G. E. Wood, N. F. J. Almeida, L. Woo, Y. Chen, I. T. Paulsen, J. A. Eisen, P. D. Karp,

D. S. Bovee, P. Chapman, J. Clendenning, G. Deatherage, W. Gillet, C. Grant, T. Kutyavin, R. Levy, M. J. Li, E. McClelland, A. Palmieri, C. Raymond, G. Rouse, C. Saenphimmachak, Z. Wu, P. Romero, D. Gordon, S. Zhang, H. Yoo, Y. Tao, P. Biddle, M. Jung, W. Krespan, M. Perry, B. Gordon-Kamm, L. Liao, S. Kim, C. Hendrick, Z. Y. Zhao, M. Dolan, F. Chumley, S. V. Tingey, J. F. Tomb, M. P. Gordon, M. V. Olson, and E. W. Nester. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science*, 294(5550):2317–23, 2001.

[122] B. Goodner, G. Hinkle, S. Gattung, N. Miller, M. Blanchard, B. Qurollo, B. S. Goldman, Y. Cao, M. Askenazi, C. Halling, L. Mullin, K. Houmiel, J. Gordon, M. Vaudin, O. Iartchouk, A. Epp, F. Liu, C. Wollam, M. Allinger, D. Doughty, C. Scott, C. Lappas, B. Markelz, C. Flanagan, C. Crowell, J. Gurson, C. Lomo, C. Sear, G. Strub, C. Cielo, and S. Slater. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science*, 294(5550):2323–8, 2001.

[123] S. E. Stachel and E. W. Nester. The genetic and transcriptional organization of the vir region of the A6 Ti plasmid of *Agrobacterium tumefaciens*. *EMBO J*, 5(7):1445–54, 1986.

[124] J. Escudero and B. Hohn. Transfer and Integration of T-DNA without Cell Injury in the Host Plant. *Plant Cell*, 9(12):2135–2142, 1997.

[125] C. A. McCullen and A. N. Binns. *Agrobacterium tumefaciens* and plant cell interactions and activities required for interkingdom macromolecular transfer. *Annu Rev Cell Dev Biol*, 22:101–27, 2006.

[126] X. Hu, J. Zhao, W. F. DeGrado, and A. N. Binns. *Agrobacterium tumefaciens* recognizes its host environment using ChvE to bind diverse plant sugars as virulence signals. *Proc Natl Acad Sci USA*, 110(2):678–83, 2013.

[127] H.-Y. Wu, P.-C. Chung, H.-W. Shih, S.-R. Wen, and E.-M. Lai. Secretome analysis uncovers an Hcp-family protein secreted via a type VI secretion system in *Agrobacterium tumefaciens*. *J Bacteriol*, 190(8):2841–50, 2008.

[128] A. Brencic and S. C. Winans. Detection of and response to signals involved in host-microbe interactions by plant-associated bacteria. *Microbiol Mol Biol Rev*, 69(1):155–94, 2005.

[129] P. J. Christie, J. E. Ward, S. C. Winans, and E. W. Nester. The *Agrobacterium tumefaciens* virE2 gene product is a single-stranded-DNA-binding protein that associates with T-DNA. *J Bacteriol*, 170(6):2659–67, 1988.

[130] E. Cascales and P. J. Christie. Definition of a bacterial type IV secretion pathway for a

DNA substrate. *Science*, 304(5674):1170–3, 2004.

[131] M. De Cleene and J. De Ley. The host range of crown gall. *The Botanical Review*, 42(4):389–466, 1976.

[132] M. A. Escobar and A. M. Dandekar. *Agrobacterium tumefaciens* as an agent of disease. *Trends Plant Sci*, 8(8):380–6, 2003.

[133] S. B. Gelvin. *Agrobacterium*-mediated plant transformation: the biology behind the ”gene-jockeying” tool. *Microbiol Mol Biol Rev*, 67(1):16–37, table of contents, 2003.

[134] B. R. Frame, H. Shou, R. K. Chikwamba, Z. Zhang, C. Xiang, T. M. Fonger, S. E. K. Pegg, B. Li, D. S. Nettleton, D. Pei, and K. Wang. *Agrobacterium tumefaciens*-mediated transformation of maize embryos using a standard binary vector system. *Plant Physiol*, 129(1):13–22, 2002.

[135] W. A. Harwood, J. G. Bartlett, S. C. Alves, M. Perry, M. A. Smedley, N. Leyland, and J. W. Snape. Barley transformation using *Agrobacterium*-mediated techniques. *Methods Mol Biol*, 478:137–47, 2009.

[136] Y. Hiei, S. Ohta, T. Komari, and T. Kumashiro. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J*, 6(2):271–82, 1994.

[137] S. Sheerman and M. W. Bevan. A rapid transformation method for *Solanum tuberosum* using binary *Agrobacterium tumefaciens* vectors. *Plant Cell Reports*, 7(1):13–16, 1988.

[138] E. Schneider, V. Eckey, D. Weidlich, N. Wiesemann, A. Vahedi-Faridi, P. Thaben, and W. Saenger. Receptor-transporter interactions of canonical ATP-binding cassette import systems in prokaryotes. *Eur J Cell Biol*, 91(4):311–7, 2012.

[139] I. Wilms, B. Voss, W. R. Hess, L. I. Leichert, and F. Narberhaus. Small RNA-mediated control of the *Agrobacterium tumefaciens* GABA binding protein. *Mol Microbiol*, 80(2):492–506, 2011.

[140] A. Overloeper, A. Kraus, R. Gurski, P. R. Wright, J. Georg, W. R. Hess, and F. Narberhaus. Two separate modules of the conserved regulatory RNA AbcR1 address multiple target mRNAs in and outside of the translation initiation region. *RNA Biol*, 11(5):624–40, 2014.

[141] R. Enns. *It’s a Nonlinear World*. Springer Undergraduate Texts in Mathematics and Technology. Springer New York, 2010.

[142] C. M. Sharma, S. Hoffmann, F. Darfeuille, J. Reignier, S. Findeiß, A. Sittka, S. Chabas, K. Reiche, J. Hackermüller, R. Reinhardt, P. F. Stadler, and J. Vogel. The primary transcriptome of the major

human pathogen *Helicobacter pylori*. *Nature*, 464(7286):250–5, 2010.

[143] N. J. Croucher and N. R. Thomson. Studying bacterial transcriptomes using RNA-seq. *Curr Opin Microbiol*, 13(5):619–24, 2010.

[144] R. Backofen and W. R. Hess. Computational prediction of sRNAs and their targets in bacteria. *RNA Biol*, 7(1):33–42, 2010.

[145] R. McClure, D. Balasubramanian, Y. Sun, M. Bobrovskyy, P. Sumby, C. A. Genco, C. K. Vanderpool, and B. Tjaden. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res*, 41(14):e140, 2013.

[146] U. Pfreundt, M. Kopf, N. Belkin, I. Berman-Frank, and W. R. Hess. The primary transcriptome of the marine diazotroph *Trichodesmium erythraeum* IMS101. *Sci Rep*, 4:6187, 2014.

[147] Y. Shimoni, G. Friedlander, G. Hetzroni, G. Niv, S. Altuvia, O. Biham, and H. Margalit. Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Syst Biol*, 3:138, 2007.

[148] S. Altuvia and E. G. Wagner. Switching on and off with RNA. *Proc Natl Acad Sci USA*, 97(18):9824–6, 2000.

[149] C. L. Beisel and G. Storz. The base-pairing RNA Spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. *Mol Cell*, 41(3):286–97, 2011.

[150] K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, and S. Gottesman. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev*, 15(13):1637–51, 2001.

[151] L. Argaman, R. Hershberg, J. Vogel, G. Bejerano, E. G. Wagner, H. Margalit, and S. Altuvia. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol*, 11(12):941–50, 2001.

[152] E. Rivas, R. J. Klein, T. A. Jones, and S. R. Eddy. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol*, 11(17):1369–73, 2001.

[153] R. Hershberg, S. Altuvia, and H. Margalit. A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res*, 31(7):1813–20, 2003.

[154] J. Vogel. A rough guide to the non-coding RNA world of *Salmonella*. *Mol Microbiol*, 71(1):1–11, 2009.

[155] J. Mitschke, J. Georg, I. Scholz, C. M. Sharma, D. Dienst, J. Bantscheff, B. Voß, C. Steglich, A. Wilde, J. Vogel, and W. R. Hess. An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci USA*, 108(5):2124–9, 2011.

[156] C. Schmidtke, S. Findeiss, C. M. Sharma, J. Kuhfuss, S. Hoffmann, J. Vogel, P. F. Stadler, and U. Bonas. Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions. *Nucleic Acids Res*, 40(5):2020–31, 2012.

[157] I. Wilms, A. Overloper, M. Nowrousian, C. M. Sharma, and F. Narberhaus. Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen *Agrobacterium tumefaciens*. *RNA Biol*, 9(4):446–57, 2012.

[158] G. Dugar, A. Herbig, K. U. Forstner, N. Heidrich, R. Reinhardt, K. Nieselt, and C. M. Sharma. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS Genet*, 9(5):e1003495, 2013.

[159] M. Kopf, S. Klaehn, I. Scholz, J. K. F. Matthiessen, W. R. Hess, and B. Voss. Comparative analysis of the primary transcriptome of Synechocystis sp. PCC 6803. *DNA Res*, 21(5):527–39, 2014.

[160] A. M. Nuss, A. K. Heroven, B. Waldmann, J. Reinkensmeier, M. Jarek, M. Beckstette, and P. Dersch. Transcriptomic profiling of *Yersinia pseudotuberculosis* reveals reprogramming of the Crp regulon by temperature and uncovers Crp as a master regulator of small RNAs. *PLoS Genet*, 11(3):e1005087, 2015.

[161] B. Fan, L. Li, Y. Chao, K. Forstner, J. Vogel, R. Borriss, and X.-Q. Wu. dRNA-Seq Reveals Genomewide TSSs and Noncoding RNAs of Plant Beneficial Rhizobacterium *Bacillus amyloliquefaciens* FZB42. *PLoS One*, 10(11):e0142002, 2015.

[162] A. M. Sass, H. Van Acker, K. U. Forstner, F. Van Nieuwerburgh, D. Deforce, J. Vogel, and T. Coenye. Genome-wide transcription start site profiling in biofilm-grown *Burkholderia cenocepacia* J2315. *BMC Genomics*, 16:775, 2015.

[163] J. Babski, K. A. Haas, D. Nather-Schindler, F. Pfeiffer, K. U. Forstner, M. Hammelmann, R. Hilker, A. Becker, C. M. Sharma, A. Marchfelder, and J. Soppa. Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics*, 17(1):629, 2016.

[164] J. Tomizawa, T. Itoh, G. Selzer, and T. Som. Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc Natl Acad Sci USA*, 78(3):1421–5, 1981.

[165] J. Georg and W. R. Hess. *cis*-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev*, 75(2):286–300, 2011.

[166] Y. He, B. Vogelstein, V. E. Velculescu, N. Papadopoulos, and K. W. Kinzler. The

antisense transcriptomes of human cells. *Science*, 322(5909):1855–7, 2008.

[167] U. Dühring, I. M. Axmann, W. R. Hess, and A. Wilde. An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proc Natl Acad Sci USA*, 103(18):7054–8, 2006.

[168] R. R. Breaker. Prospects for riboswitch discovery and analysis. *Mol Cell*, 43(6):867–79, 2011.

[169] A. Serganov and E. Nudler. A decade of riboswitches. *Cell*, 152(1-2):17–24, 2013.

[170] J. Shine and L. Dalgarno. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci USA*, 71(4):1342–6, 1974.

[171] A. Nocker, T. Hausherr, S. Balsiger, N. P. Krstulovic, H. Hennecke, and F. Narberhaus. A mRNA-based thermosensor controls expression of rhizobial heat shock genes. *Nucleic Acids Res*, 29(23):4800–7, 2001.

[172] J. Kortmann, S. Sczodrok, J. Rinnenthal, H. Schwalbe, and F. Narberhaus. Translation on demand by a simple RNA-based thermosensor. *Nucleic Acids Res*, 39(7):2855–68, 2011.

[173] F. Righetti, A. M. Nuss, C. Twittenhoff, S. Beele, K. Urban, S. Will, S. H. Bernhart, P. F. Stadler, P. Dersch, and F. Narberhaus. Temperature-responsive in vitro RNA structurome of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA*, 2016.

[174] L. He and G. J. Hannon. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*, 5(7):522–31, 2004.

[175] H. Guo, N. T. Ingolia, J. S. Weissman, and D. P. Bartel. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308):835–40, 2010.

[176] A. Busch, A. S. Richter, and R. Backofen. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24(24):2849–56, 2008.

[177] M. R. Fabian, N. Sonenberg, and W. Filipowicz. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem*, 79:351–79, 2010.

[178] J. T. Cuperus, N. Fahlgren, and J. C. Carrington. Evolution and functional diversification of MIRNA genes. *Plant Cell*, 23(2):431–42, 2011.

[179] J. Vogel and B. F. Luisi. Hfq and its constellation of RNA. *Nat Rev Microbiol*, 9(8):578–89, 2011.

[180] K. I. Udekwu. Transcriptional and post-transcriptional regulation of the Escherichia coli luxS mRNA; involvement of the sRNA MicA. *PLoS One*, 5(10):e13449, 2010.

[181] D. Jäger, S. R. Pernitzsch, A. S. Richter, R. Backofen, C. M. Sharma, and R. A.

Schmitz. An archaeal sRNA targeting *cis*- and *trans*- encoded mRNAs via two distinct domains. *Nucleic Acids Res*, 40(21):10964–79, 2012.

[182] Y. Chao, K. Papenfort, R. Reinhardt, C. M. Sharma, and J. Vogel. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J*, 31(20):4005–19, 2012.

[183] C. P. Corcoran, D. Podkaminski, K. Papenfort, J. H. Urban, J. C. D. Hinton, and J. Vogel. Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol*, 84(3):428–45, 2012.

[184] M. S. Guo, T. B. Updegrove, E. B. Gogol, S. A. Shabalina, C. A. Gross, and G. Storz. MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev*, 28(14):1620–34, 2014.

[185] M. Bobrovskyy and C. K. Vanderpool. Diverse mechanisms of post-transcriptional repression by the small RNA regulator of glucose-phosphate stress. *Mol Microbiol*, 99(2):254–73, 2016.

[186] J. Babski, L.-K. Maier, R. Heyer, K. Jaschinski, D. Prasse, D. Jager, L. Randau, R. A. Schmitz, A. Marchfelder, and J. Soppa. Small regulatory RNAs in Archaea. *RNA Biol*, 11(5):484–93, 2014.

[187] D. Beyer, E. Skripkin, J. Wadzack, and K. H. Nierhaus. How the ribosome moves along the mRNA during protein synthesis. *J Biol Chem*, 269(48):30713–7, 1994.

[188] F. Darfeuille, C. Unoson, J. Vogel, and E. G. H. Wagner. An antisense RNA inhibits translation by competing with standby ribosomes. *Mol Cell*, 26(3):381–92, 2007.

[189] E. G. H. Wagner and C. Unoson. The toxin-antitoxin system tisB-istR1: Expression, regulation, and biological role in persister phenotypes. *RNA Biol*, 9(12):1513–9, 2012.

[190] Q. Yang, N. Figueroa-Bossi, and L. Bossi. Translation enhancing ACA motifs and their silencing by a bacterial small regulatory RNA. *PLoS Genet*, 10(1):e1004026, 2014.

[191] B. Večerek, I. Moll, and U. Bläsi. Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. *EMBO J*, 26(4):965–75, 2007.

[192] K. J. Bandyra, N. Said, V. Pfeiffer, M. W. Gorna, J. Vogel, and B. F. Luisi. The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. *Mol Cell*, 2012.

[193] E. Masse, F. E. Escorcia, and S. Gottesman. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev*, 17(19):2374–83, 2003.

[194] E. Morfeldt, D. Taylor, A. von Gabain, and S. Arvidson. Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J*, 14(18):4569–77, 1995.

[195] K. S. Fröhlich and J. Vogel. Activation of gene expression by small RNA. *Curr Opin Microbiol*, 12(6):674–82, 2009.

[196] T. Soper, P. Mandin, N. Majdalani, S. Gottesman, and S. A. Woodson. Positive regulation by small RNAs and the role of Hfq. *Proc Natl Acad Sci USA*, 107(21):9602–7, 2010.

[197] K. Papenfort, Y. Sun, M. Miyakoshi, C. K. Vanderpool, and J. Vogel. Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell*, 153(2):426–37, 2013.

[198] K. S. Frohlich, K. Papenfort, A. Fekete, and J. Vogel. A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J*, 32(22):2963–79, 2013.

[199] M. Y. Liu, G. Gui, B. Wei, J. F. r. Preston, L. Oakford, U. Yuksel, D. P. Giedroc, and T. Romeo. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem*, 272(28):17502–10, 1997.

[200] T. Romeo, C. A. Vakulskas, and P. Babitzke. Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environ Microbiol*, 15(2):313–24, 2013.

[201] E. Holmqvist, P. R. Wright, L. Li, T. Bischler, L. Barquist, R. Reinhardt, R. Backofen, and J. Vogel. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J*, 2016.

[202] B. Steuten, S. Schneider, and R. Wagner. 6s RNA: recent answers–future questions. *Mol Microbiol*, 91(4):641–8, 2014.

[203] K. M. Wassarman and G. Storz. 6S RNA regulates *E. coli* RNA polymerase activity. *Cell*, 101(6):613–23, 2000.

[204] A. E. Trotochaud and K. M. Wassarman. 6s RNA function enhances long-term cell survival. *J Bacteriol*, 186(15):4978–85, 2004.

[205] K. M. Wassarman. 6S RNA: a small RNA regulator of transcription. *Curr Opin Microbiol*, 10(2):164–8, 2007.

[206] M. Miyakoshi, Y. Chao, and J. Vogel. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J*, 34(11):1478–92, 2015.

[207] J. J. Tree, S. Granneman, S. P. McAteer, D. Tollervey, and D. L. Gally. Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*. *Mol Cell*, 55(2):199–213, 2014.

[208] S. Melamed, A. Peer, R. Faigenbaum-Romm, Y. E. Gatt, N. Reiss, A. Bar, Y. Altuvia, L. Argaman, and H. Margalit.

Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol Cell*, 63(5):884–97, 2016.

[209] C. S. Wadler and C. K. Vanderpool. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc Natl Acad Sci USA*, 104(51):20454–9, 2007.

[210] G. Storz, Y. I. Wolf, and K. S. Ramamurthi. Small proteins can no longer be ignored. *Annu Rev Biochem*, 83:753–77, 2014.

[211] D. Prasse, J. Thomsen, R. De Santis, J. Muntel, D. Becher, and R. A. Schmitz. First description of small proteins encoded by spRNAs in *Methanosarcina mazei* strain Go1. *Biochimie*, 117:138–48, 2015.

[212] D. Baumgartner, M. Kopf, S. Klähn, C. Steglich, and W. R. Hess. Small proteins in cyanobacteria provide a paradigm for the functional analysis of the bacterial micro-proteome. *BMC Microbiology*, 16(1):285, 2016.

[213] K. Neuhaus, R. Landstorfer, S. Simon, S. Schober, P. R. Wright, C. Smith, R. Backofen, R. Wecko, D. A. Keim, and S. Scherer. Differentiation of ncRNAs from small mRNAs in *Escherichia coli* O157:H7 EDL933 (EHEC) by combined RNAseq and RIBOseq - ryhB encodes the regulatory RNA RyhB and a peptide, RyhP. *Submitted*, 2016.

[214] K. Papenfort and J. Vogel. Regulatory RNA in bacterial pathogens. *Cell Host Microbe*, 8(1):116–27, 2010.

[215] J. P. Bardill and B. K. Hammer. Non-coding sRNAs regulate virulence in the bacterial pathogen *Vibrio cholerae*. *RNA Biol*, 9(4):392–401, 2012.

[216] S. C. Pulvermacher, L. T. Stauffer, and G. V. Stauffer. Role of the sRNA GcvB in regulation of cycA in *Escherichia coli*. *Microbiology*, 155(Pt 1):106–14, 2009.

[217] S. C. Pulvermacher, L. T. Stauffer, and G. V. Stauffer. The small RNA GcvB regulates *sstT* mRNA expression in *Escherichia coli*. *J Bacteriol*, 191(1):238–48, 2009.

[218] T. Møller, T. Franch, C. Udesen, K. Gerdes, and P. Valentin-Hansen. Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev*, 16(13):1696–706, 2002.

[219] C. L. Beisel, T. B. Updegrove, B. J. Janson, and G. Storz. Multiple factors dictate target selection by Hfq-binding small RNAs. *EMBO J*, 31(8):1961–74, 2012.

[220] H. Salvail, P. Lanthier-Bourbonnais, J. M. Sobota, M. Caza, J.-A. M. Benjamin, M. E. S. Mendieta, F. Lépine, C. M. Dozois, J. Imlay, and E. Massé. A small RNA promotes siderophore production through transcriptional and metabolic remodeling. *Proc Natl Acad Sci USA*, 107(34):15223–8, 2010.

[221] R. A. Lease, M. E. Cusick, and M. Belfort. Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc Natl Acad Sci USA*, 95(21):12456–61, 1998.

[222] N. Majdalani, D. Hernandez, and S. Gottesman. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol Microbiol*, 46(3):813–26, 2002.

[223] E. Holmqvist, C. Unoson, J. Reimegard, and E. G. H. Wagner. A mixed double negative feedback loop between the sRNA MicF and the global regulator Lrp. *Mol Microbiol*, 84(3):414–27, 2012.

[224] N. Sedlyarova, I. Shamovsky, B. K. Bharati, V. Epshtein, J. Chen, S. Gottesman, R. Schroeder, and E. Nudler. sRNA-Mediated Control of Transcription Termination in E. coli. *Cell*, 167(1):111–121.e13, 2016.

[225] A. Grylak-Mielnicka, V. Bidnenko, J. Bardowski, and E. Bidnenko. Transcription termination factor Rho: a hub linking diverse physiological processes in bacteria. *Microbiology*, 162(3):433–47, 2016.

[226] P. R. Wright, J. Georg, M. Mann, D. A. Sorescu, A. S. Richter, S. Lott, R. Kleinkauf, W. R. Hess, and R. Backofen. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res*, 42(Web Server issue):W119–23, 2014. PRW, JG and MM contributed equally to this work.

[227] Y. Takeda and H. Avila. Structure and gene expression of the E. coli Mn-superoxide dismutase gene. *Nucleic Acids Res*, 14(11):4577–89, 1986.

[228] J. A. Imlay. Pathways of oxidative damage. *Annu Rev Microbiol*, 57:395–418, 2003.

[229] K. Salmon, S.-p. Hung, K. Mekjian, P. Baldi, G. W. Hatfield, and R. P. Gunsalus. Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *J Biol Chem*, 278(32):29837–55, 2003.

[230] B. A. Lazazzera, H. Beinert, N. Khoroshilova, M. C. Kennedy, and P. J. Kiley. DNA binding and dimerization of the Fe-S-containing FNR protein from *Escherichia coli* are regulated by oxygen. *J Biol Chem*, 271(5):2762–8, 1996.

[231] Y. Kang, K. D. Weber, Y. Qiu, P. J. Kiley, and F. R. Blattner. Genome-wide expression analysis indicates that FNR of *Escherichia coli* K-12 regulates a large number of genes of unknown function. *J Bacteriol*, 187(3):1135–60, 2005.

[232] C. Constantinidou, J. L. Hobman, L. Griffiths, M. D. Patel, C. W. Penn, J. A. Cole, and T. W. Overton. A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL,

and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. *J Biol Chem*, 281(8):4802–15, 2006.

[233] A. Boysen, J. Moller-Jensen, B. Kallipolitis, P. Valentin-Hansen, and M. Overgaard. Translational regulation of gene expression by an anaerobically induced small noncoding RNA in *Escherichia coli*. *J Biol Chem*, 285(14):10690–702, 2010.

[234] S. Durand and G. Storz. Reprogramming of anaerobic metabolism by the FnrS small RNA. *Mol Microbiol*, 75(5):1215–31, 2010.

[235] E. P. Nawrocki, S. W. Burge, A. Bateman, J. Daub, R. Y. Eberhardt, S. R. Eddy, E. W. Floden, P. P. Gardner, T. A. Jones, J. Tate, and R. D. Finn. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res*, 43(Database issue):D130–7, 2015.

[236] A. Sittka, S. Lucchini, K. Papenfort, C. M. Sharma, K. Rolle, T. T. Binnewies, J. C. D. Hinton, and J. Vogel. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet*, 4(8):e1000163, 2008.

[237] M. K. SANDS and R. B. ROBERTS. The effects of a tryptophan-histidine deficiency in a mutant of *Escherichia coli*. *J Bacteriol*, 63(4):505–11, 1952.

[238] M. L. Urbanowski, L. T. Stauffer, and G. V. Stauffer. The gcvB gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol Microbiol*, 37(4):856–68, 2000.

[239] M. G. Jorgensen, J. S. Nielsen, A. Boysen, T. Franch, J. Moller-Jensen, and P. Valentin-Hansen. Small regulatory RNAs control the multi-cellular adhesive lifestyle of *Escherichia coli*. *Mol Microbiol*, 84(1):36–50, 2012.

[240] A. Coornaert, C. Chiaruttini, M. Springer, and M. Guillier. Post-transcriptional control of the *Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. *PLoS Genet*, 9(1):e1003156, 2013.

[241] J. Vogel, V. Bartels, T. H. Tang, G. Churakov, J. G. Slagter-Jager, A. Huttenhofer, and E. G. H. Wagner. RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res*, 31(22):6435–43, 2003.

[242] S. C. Andrews, A. K. Robinson, and F. Rodriguez-Quinones. Bacterial iron homeostasis. *FEMS Microbiol Rev*, 27(2-3):215–37, 2003.

[243] K. Hantke. Cloning of the repressor protein gene of iron-regulated systems in *Escherichia coli* K12. *Mol Gen Genet*, 197(2):337–41, 1984.

[244] K. Hantke. Iron and metal regulation in

bacteria. *Curr Opin Microbiol*, 4(2):172–7, 2001.

[245] L. Escolar, J. Perez-Martin, and V. de Lorenzo. Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol*, 181(20):6223–9, 1999.

[246] M. J. Gruer and J. R. Guest. Two genetically-distinct and differentially-regulated aconitases (AcnA and AcnB) in *Escherichia coli*. *Microbiology*, 140 (Pt 10):2531–41, 1994.

[247] S. Dubrac and D. Touati. Fur positive regulation of iron superoxide dismutase in *Escherichia coli*: functional analysis of the sodB promoter. *J Bacteriol*, 182(13):3802–8, 2000.

[248] E. Massé and S. Gottesman. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA*, 99(7):4620–5, 2002.

[249] D. H. Flint, M. H. Emptage, and J. R. Guest. Fumarase a from *Escherichia coli*: purification and characterization as an iron-sulfur cluster containing enzyme. *Biochemistry*, 31(42):10331–7, 1992.

[250] C. K. Vance and A. F. Miller. Novel insights into the basis for *Escherichia coli* superoxide dismutase's metal ion specificity from Mn-substituted FeSOD and its very high E(m). *Biochemistry*, 40(43):13079–87, 2001.

[251] B. Vecerek, I. Moll, T. Afonyushkin, V. Kaberdin, and U. Blasi. Interaction of the RNA chaperone Hfq with mRNAs: direct and indirect roles of Hfq in iron metabolism of *Escherichia coli*. *Mol Microbiol*, 50(3):897–909, 2003.

[252] T. A. Geissmann and D. Touati. Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J*, 23(2):396–405, 2004.

[253] E. Masse, C. K. Vanderpool, and S. Gottesman. Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol*, 187(20):6962–71, 2005.

[254] K. Prévost, H. Salvail, G. Desnoyers, J.-F. Jacques, É. Phaneuf, and E. Massé. The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Mol Microbiol*, 64(5):1260–73, 2007.

[255] G. Desnoyers, A. Morissette, K. Prévost, and E. Massé. Small RNA-induced differential degradation of the polycistronic mRNA *iscRSUA*. *EMBO J*, 28(11):1551–61, 2009.

[256] K. Prévost, G. Desnoyers, J.-F. Jacques, F. Lavoie, and E. Massé. Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes Dev*, 25(4):385–96, 2011.

[257] J. Bos, Y. Duverger, B. Thouvenot, C. Chiaruttini, C. Branlant, M. Springer, B. Charpentier, and F. Barras. The sRNA RyhB regulates the synthesis of the *Escherichia coli* methionine sulfoxide reductase MsrB but not MsrA. *PLoS One*, 8(5):e63647, 2013.

[258] J.-A. M. Benjamin and E. Masse. The iron-sensing aconitase B binds its own mRNA to prevent sRNA-induced mRNA cleavage. *Nucleic Acids Res*, 42(15):10023–36, 2014.

[259] L. Argaman, M. Elgrably-Weiss, T. Hershko, J. Vogel, and S. Altuvia. RelA protein stimulates the activity of RyhB small RNA by acting on RNA-binding protein Hfq. *Proc Natl Acad Sci USA*, 109(12):4621–6, 2012.

[260] J. Wang, W. Rennie, C. Liu, C. S. Carmack, K. Prevost, M.-P. Caron, E. Masse, Y. Ding, and J. T. Wade. Identification of bacterial sRNA regulatory targets using ribosome profiling. *Nucleic Acids Res*, 43(21):10308–20, 2015.

[261] R. Balbontin, N. Villagra, M. Pardos de la Gandara, G. Mora, N. Figueroa-Bossi, and L. Bossi. Expression of IroN, the salmochelin siderophore receptor, requires mRNA activation by RyhB small RNA homologues. *Mol Microbiol*, 100(1):139–55, 2016.

[262] M.-C. Carrier, D. Lalaouna, and E. Masse. A game of tag: MAPS catches up on RNA interactomes. *RNA Biol*, 13(5):473–6, 2016.

[263] P. J. Wilderman, N. A. Sowa, D. J. FitzGerald, P. C. FitzGerald, S. Gottesman, U. A. Ochsner, and M. L. Vasil. Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc Natl Acad Sci USA*, 101(26):9792–7, 2004.

[264] A. Gaballa, H. Antelmann, C. Aguilar, S. K. Khakh, K.-B. Song, G. T. Smaldone, and J. D. Helmann. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc Natl Acad Sci USA*, 105(33):11927–32, 2008.

[265] G. T. Smaldone, H. Antelmann, A. Gaballa, and J. D. Helmann. The FsrA sRNA and FbpB protein mediate the iron-dependent induction of the *Bacillus subtilis* lutABC iron-sulfur-containing oxidases. *J Bacteriol*, 194(10):2586–93, 2012.

[266] B. Gorke and J. Stulke. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*, 6(8):613–24, 2008.

[267] M. Sumiya, E. O. Davis, L. C. Packman, T. P. McDonald, and P. J. Henderson. Molecular genetics of a receptor protein for D-xylose, encoded by the gene xylF, in *Escherichia coli*. *Receptors Channels*, 3(2):117–28, 1995.

[268] R. W. Hogg, C. Voelker, and I. Von Carlowitz. Nucleotide sequence and analysis of the mgl operon of *Escherichia coli* K12. *Mol Gen Genet*, 229(3):453–9, 1991.

[269] D. A. Polayes, P. W. Rice, M. M. Garner, and J. E. Dahlberg. Cyclic AMP-cyclic AMP receptor protein as a repressor of transcription of the spf gene of *Escherichia coli*. *J Bacteriol*, 170(7):3110–4, 1988.

[270] I. Pastan and R. Perlman. Cyclic adenosine monophosphate in bacteria. *Science*, 169(3943):339–44, 1970.

[271] H. Ishizuka, A. Hanamura, T. Inada, and H. Aiba. Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the crp gene. *EMBO J*, 13(13):3077–82, 1994.

[272] B. G. Sahagan and J. E. Dahlberg. A small, unstable RNA molecule of *Escherichia coli*: spot 42 RNA. II. Accumulation and distribution. *J Mol Biol*, 131(3):593–605, 1979.

[273] T. Moller, T. Franch, P. Hojrup, D. R. Keene, H. P. Bachinger, R. G. Brennan, and P. Valentin-Hansen. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell*, 9(1):23–30, 2002.

[274] B. G. Sahagan and J. E. Dahlberg. A small, unstable RNA molecule of *Escherichia coli*: spot 42 RNA. I. Nucleotide sequence analysis. *J Mol Biol*, 131(3):573–92, 1979.

[275] U. Alon. Network motifs: theory and experimental approaches. *Nat Rev Genet*, 8(6):450–61, 2007.

[276] G. Desnoyers and E. Masse. Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes Dev*, 26(7):726–39, 2012.

[277] G. A. Hansen, R. Ahmad, E. Hjerde, C. G. Fenton, N.-P. Willassen, and P. Haugen. Expression profiling reveals Spot 42 small RNA as a key regulator in the central metabolism of *Aliivibrio salmonicida*. *BMC Genomics*, 13:37, 2012.

[278] T. R. Cech and J. A. Steitz. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, 157(1):77–94, 2014.

[279] T. Glisovic, J. L. Bachorik, J. Yong, and G. Dreyfuss. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett*, 582(14):1977–86, 2008.

[280] A. G. Baltz, M. Munschauer, B. Schwanhausser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich, and M. Landthaler. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell*, 46(5):674–90, 2012.

[281] A. Castello, B. Fischer, K. Eichelbaum, R. Horos, B. M. Beckmann, C. Strein, N. E. Davey, D. T. Humphreys, T. Preiss,

L. M. Steinmetz, J. Krijgsveld, and M. W. Hentze. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 149(6):1393–406, 2012.

[282] D. Maticzka, S. J. Lange, F. Costa, and R. Backofen. GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biol*, 15(1):R17, 2014.

[283] J. M. Taliaferro, N. J. Lambert, P. H. Sudmant, D. Dominguez, J. J. Merkin, M. S. Alexis, C. A. Bazile, and C. B. Burge. RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. *Mol Cell*, 2016.

[284] X.-D. Fu and M. J. Ares. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet*, 15(10):689–701, 2014.

[285] E. Elkayam, C.-D. Kuhn, A. Tocilj, A. D. Haase, E. M. Greene, G. J. Hannon, and L. Joshua-Tor. The Structure of Human Argonaute-2 in Complex with miR-20a. *Cell*, 150(1):100–10, 2012.

[286] C. M. Brennan and J. A. Steitz. HuR and mRNA stability. *Cell Mol Life Sci*, 58(2):266–77, 2001.

[287] D. D. Licatalosi, A. Mele, J. J. Fak, J. Ule, M. Kayikci, S. W. Chi, T. A. Clark, A. C. Schweitzer, J. E. Blume, X. Wang, J. C. Darnell, and R. B. Darnell. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, 456(7221):464–9, 2008.

[288] J. Konig, K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe, and J. Ule. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*, 17(7):909–15, 2010.

[289] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. J. Ascano, A.-C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan, and T. Tuschl. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141(1):129–41, 2010.

[290] M. Uhl, T. Houwaart, G. Corrado, P. R. Wright, and R. Backofen. Computational Analysis of CLIP-seq Data. *Submitted*, 2016.

[291] M. T. Franze de Fernandez, L. Eoyang, and J. T. August. Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature*, 219(5154):588–90, 1968.

[292] T. B. Updegrove, A. Zhang, and G. Storz. Hfq: the flexible RNA matchmaker. *Curr Opin Microbiol*, 30:133–8, 2016.

[293] C. Sauter, J. Basquin, and D. Suck. Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. *Nucleic Acids Res*, 31(14):4091–8, 2003.

[294] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N.

Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res*, 28(1):235–42, 2000.

[295] P. W. Rose, A. Prlic, C. Bi, W. F. Bluhm, C. H. Christie, S. Dutta, R. K. Green, D. S. Goodsell, J. D. Westbrook, J. Woo, J. Young, C. Zardecki, H. M. Berman, P. E. Bourne, and S. K. Burley. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res*, 43(Database issue):D345–56, 2015.

[296] M. Biasini. pv: v1.8.1, 2015.

[297] D. Dienst, U. Duhring, H.-J. Mollenkopf, J. Vogel, J. Golecki, W. R. Hess, and A. Wilde. The cyanobacterial homologue of the RNA chaperone Hfq is essential for motility of *Synechocystis* sp. PCC 6803. *Microbiology*, 154(Pt 10):3134–43, 2008.

[298] H. C. Tsui, H. C. Leung, and M. E. Winkler. Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. *Mol Microbiol*, 13(1):35–49, 1994.

[299] G. T. Robertson and R. M. J. Roop. The *Brucella abortus* host factor I (HF-I) protein contributes to stress resistance during stationary phase and is a major determinant of virulence in mice. *Mol Microbiol*, 34(4):690–700, 1999.

[300] N. J. Tobias, A. K. Heinrich, H. Eresmann, P. R. Wright, N. Neubacher, R. Backofen, and H. B. Bode. *Photorhabdus*-nematode symbiosis is dependent on *hfq*-mediated regulation of secondary metabolites. *Environ Microbiol*, 2016.

[301] C. Bohn, C. Rigoulay, and P. Bouloc. No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. *BMC Microbiol*, 7:10, 2007.

[302] T. M. Link, P. Valentin-Hansen, and R. G. Brennan. Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci USA*, 2009.

[303] M. A. Schumacher, R. F. Pearson, T. Moller, P. Valentin-Hansen, and R. G. Brennan. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J*, 21(13):3546–56, 2002.

[304] A. Nikulin, E. Stolboushkina, A. Perederina, I. Vassilieva, U. Blaesi, I. Moll, G. Kachalova, S. Yokoyama, D. Vassylyev, M. Garber, and S. Nikonov. Structure of *Pseudomonas aeruginosa* Hfq protein. *Acta Crystallogr D Biol Crystallogr*, 61(Pt 2):141–6, 2005.

[305] N. Horstmann, J. Orans, P. Valentin-Hansen, S. A. r. Shelburne, and R. G. Brennan. Structural mechanism of *Staphylococcus aureus* Hfq binding to an RNA A-tract. *Nucleic Acids Res*, 40(21):11023–35, 2012.

[306] J. S. Nielsen, A. Boggild, C. B. F. Andersen, G. Nielsen, A. Boysen, D. E. Broder-

sen, and P. Valentin-Hansen. An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from *Methanococcus jannaschii*. *RNA*, 13(12):2213–23, 2007.

[307] E. Sauer and O. Weichenrieder. Structural basis for RNA 3'-end recognition by Hfq. *Proc Natl Acad Sci USA*, 108(32):13065–70, 2011.

[308] T. Someya, S. Baba, M. Fujimoto, G. Kawai, T. Kumasaka, and K. Nakamura. Crystal structure of Hfq from *Bacillus subtilis* in complex with SELEX-derived RNA aptamer: insight into RNA-binding properties of bacterial Hfq. *Nucleic Acids Res*, 40(4):1856–67, 2012.

[309] A. Boggild, M. Overgaard, P. Valentin-Hansen, and D. E. Brodersen. Cyanobacteria contain a structural homologue of the Hfq protein with altered RNA-binding properties. *FEBS J*, 276(14):3904–15, 2009.

[310] E. M. Tan and H. G. Kunkel. Characteristics of a soluble nuclear antigen precipitating with sera of patients with systemic lupus erythematosus. *J Immunol*, 96(3):464–71, 1966.

[311] W. H. Reeves, S. Narain, and M. Satoh. Henry Kunkel, Stephanie Smith, clinical immunology, and split genes. *Lupus*, 12(3):213–7, 2003.

[312] S. Tharun. Roles of eukaryotic Lsm proteins in the regulation of mRNA function. *Int Rev Cell Mol Biol*, 272:149–89, 2009.

[313] P. Valentin-Hansen, M. Eriksen, and C. Udesen. The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol Microbiol*, 51(6):1525–33, 2004.

[314] C. J. Wilusz and J. Wilusz. Eukaryotic Lsm proteins: lessons from bacteria. *Nat Struct Mol Biol*, 12(12):1031–6, 2005.

[315] W. R. Hess, B. A. Berghoff, A. Wilde, C. Steglich, and G. Klug. Riboregulators and the role of Hfq in photosynthetic bacteria. *RNA Biol*, 11(5):413–26, 2014.

[316] N. Schuergers, U. Ruppert, S. Watanabe, D. J. Nurnberg, G. Lochnit, D. Dienst, C. W. Mullineaux, and A. Wilde. Binding of the RNA chaperone Hfq to the type IV pilus base is crucial for its function in *Synechocystis* sp. PCC 6803. *Mol Microbiol*, 92(4):840–52, 2014.

[317] T. Morita, K. Maki, and H. Aiba. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev*, 19(18):2176–86, 2005.

[318] V. Pfeiffer, K. Papenfort, S. Lucchini, J. C. D. Hinton, and J. Vogel. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol*, 16(8):840–6, 2009.

[319] Y. Ikeda, M. Yagi, T. Morita, and H. Aiba. Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in *Escherichia coli*. *Mol Microbiol*, 79(2):419–32, 2011.

[320] B. K. Mohanty, V. F. Maples, and S. R. Kushner. The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*. *Mol Microbiol*, 54(4):905–20, 2004.

[321] N. N. Salim and A. L. Feig. An upstream Hfq binding site in the fhlA mRNA leader region facilitates the OxyS-fhlA interaction. *PLoS One*, 5(9), 2010.

[322] A. Fender, J. Elf, K. Hampel, B. Zimmermann, and E. G. H. Wagner. RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev*, 24(23):2621–6, 2010.

[323] M. Olejniczak. Despite similar binding to the Hfq protein regulatory RNAs widely differ in their competition performance. *Biochemistry*, 50(21):4427–40, 2011.

[324] R. Hussein and H. N. Lim. Disruption of small RNA signaling caused by competition for Hfq. *Proc Natl Acad Sci USA*, 108(3):1110–5, 2011.

[325] K. I. Udekwu, F. Darfeuille, J. Vogel, J. Reimegård, E. Holmqvist, and E. G. H. Wagner. Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev*, 19(19):2355–66, 2005.

[326] E. Holmqvist, J. Reimegård, M. Sterk, N. Grantcharova, U. Römling, and E. G. H. Wagner. Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J*, 29(11):1840–50, 2010.

[327] J. B. Rice, D. Balasubramanian, and C. K. Vanderpool. Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA. *Proc Natl Acad Sci USA*, 109(40):E2691–8, 2012.

[328] Y. Chao and J. Vogel. A 3’ UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. *Mol Cell*, 61(3):352–63, 2016.

[329] K. Papenfort, V. Pfeiffer, F. Mika, S. Lucchini, J. C. D. Hinton, and J. Vogel. SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. *Mol Microbiol*, 62(6):1674–88, 2006.

[330] E. G. H. Wagner. Cycling of RNAs on Hfq. *RNA Biol*, 10(4):619–26, 2013.

[331] E. Sauer, S. Schmidt, and O. Weichenrieder. Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc Natl Acad Sci USA*, 109(24):9396–401, 2012.

[332] D. J. Schu, A. Zhang, S. Gottesman, and G. Storz. Alternative Hfq-sRNA interaction modes dictate alternative mRNA

recognition. *EMBO J*, 34(20):2557–73, 2015.

[333] B. Vecerek, L. Rajkowitsch, E. Sonnleitner, R. Schroeder, and U. Blasi. The C-terminal domain of *Escherichia coli* Hfq is required for regulation. *Nucleic Acids Res*, 36(1):133–43, 2008.

[334] D. Schilling and U. Gerischer. The *Acinetobacter baylyi* Hfq gene encodes a large protein with an unusual C terminus. *J Bacteriol*, 191(17):5553–62, 2009.

[335] M. Beich-Frandsen, B. Vecerek, P. V. Konarev, B. Sjoblom, K. Kloiber, H. Hammerle, L. Rajkowitsch, A. J. Miles, G. Kontaxis, B. A. Wallace, D. I. Svergun, R. Konrat, U. Blasi, and K. Djinovic-Carugo. Structural insights into the dynamics and function of the C-terminus of the E. coli RNA chaperone Hfq. *Nucleic Acids Res*, 39(11):4900–15, 2011.

[336] A. S. Attia, J. L. Sedillo, W. Wang, W. Liu, C. A. Brautigam, W. Winkler, and E. J. Hansen. *Moraxella catarrhalis* expresses an unusual Hfq protein. *Infect Immun*, 76(6):2520–30, 2008.

[337] A. S. Olsen, J. Moller-Jensen, R. G. Brennan, and P. Valentin-Hansen. C-terminally truncated derivatives of *Escherichia coli* Hfq are proficient in riboregulation. *J Mol Biol*, 404(2):173–82, 2010.

[338] N. N. Salim, M. A. Faner, J. A. Philip, and A. L. Feig. Requirement of upstream Hfq-binding (ARN)x elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY. *Nucleic Acids Res*, 40(16):8021–32, 2012.

[339] A. Santiago-Frangos, K. Kavita, D. J. Schu, S. Gottesman, and S. A. Woodson. C-terminal domain of the RNA chaperone Hfq drives sRNA competition and release of target RNA. *Proc Natl Acad Sci USA*, 2016.

[340] A. Zhang, K. M. Wassarman, C. Rosenow, B. C. Tjaden, G. Storz, and S. Gottesman. Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol*, 50(4):1111–24, 2003.

[341] B. A. Berghoff, J. Glaeser, C. M. Sharma, M. Zobawa, F. Lottspeich, J. Vogel, and G. Klug. Contribution of Hfq to photooxidative stress resistance and global regulation in *Rhodobacter sphaeroides*. *Mol Microbiol*, 80(6):1479–95, 2011.

[342] A. K. Dubey, C. S. Baker, K. Suzuki, A. D. Jones, P. Pandit, T. Romeo, and P. Babitzke. CsrA regulates translation of the *Escherichia coli* carbon starvation gene, cstA, by blocking ribosome access to the cstA transcript. *J Bacteriol*, 185(15):4450–60, 2003.

[343] J. Timmermans and L. Van Melderen. Post-transcriptional global regulation by CsrA in bacteria. *Cell Mol Life Sci*, 67(17):2897–908, 2010.

[344] D. White, M. E. Hart, and T. Romeo. Phylogenetic distribution of the global regula-

tory gene csrA among eubacteria. *Gene*, 182(1-2):221–3, 1996.

[345] R. D. Finn, P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate, and A. Bateman. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, 44(D1):D279–85, 2016.

[346] T. Romeo, M. Gong, M. Y. Liu, and A. M. Brun-Zinkernagel. Identification and molecular characterization of csrA, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J Bacteriol*, 175(15):4744–55, 1993.

[347] S. D. Lawhon, J. G. Frye, M. Suyemoto, S. Porwollik, M. McClelland, and C. Altier. Global regulation by CsrA in *Salmonella typhimurium*. *Mol Microbiol*, 48(6):1633–45, 2003.

[348] A. Chatterjee, Y. Cui, Y. Liu, C. K. Dumenyo, and A. K. Chatterjee. Inactivation of rsmA leads to overproduction of extracellular pectinases, cellulases, and proteases in *Erwinia carotovora* subsp. *carotovora* in the absence of the starvation/cell density-sensing signal, N-(3-oxohexanoyl)-L-homoserine lactone. *Appl Environ Microbiol*, 61(5):1959–67, 1995.

[349] A. Mukherjee, Y. Cui, Y. Liu, C. K. Dumenyo, and A. K. Chatterjee. Global regulation in *Erwinia* species by *Erwinia carotovora* rsmA, a homologue of *Escherichia coli* csrA: repression of secondary metabolites, pathogenicity and hypersensitive reaction. *Microbiology*, 142 (Pt 2):427–34, 1996.

[350] K. Lapouge, E. Sineva, M. Lindell, K. Starke, C. S. Baker, P. Babitzke, and D. Haas. Mechanism of hcnA mRNA recognition in the Gac/Rsm signal transduction pathway of *Pseudomonas fluorescens*. *Mol Microbiol*, 66(2):341–56, 2007.

[351] Y. Irie, M. Starkey, A. N. Edwards, D. J. Wozniak, T. Romeo, and M. R. Parsek. *Pseudomonas aeruginosa* biofilm matrix polysaccharide Psl is regulated transcriptionally by RpoS and post-transcriptionally by RsmA. *Mol Microbiol*, 78(1):158–72, 2010.

[352] H. Yakhnin, C. S. Baker, I. Berezin, M. A. Evangelista, A. Rassin, T. Romeo, and P. Babitzke. CsrA represses translation of sdiA, which encodes the N-acylhomoserine-L-lactone receptor of *Escherichia coli*, by binding exclusively within the coding region of sdiA mRNA. *J Bacteriol*, 193(22):6162–70, 2011.

[353] R. Stoltenburg, C. Reinemann, and B. Strehlitz. SELEX–a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng*, 24(4):381–403, 2007.

[354] A. K. Dubey, C. S. Baker, T. Romeo, and P. Babitzke. RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA*, 11(10):1579–87, 2005.

[355] C. S. Baker, I. Morozov, K. Suzuki, T. Romeo, and P. Babitzke. CsrA regulates glycogen biosynthesis by preventing translation of glgC in *Escherichia coli*. *Mol Microbiol*, 44(6):1599–610, 2002.

[356] A. Pannuri, H. Yakhnin, C. A. Vakulskas, A. N. Edwards, P. Babitzke, and T. Romeo. Translational repression of NhaR, a novel pathway for multi-tier regulation of biofilm circuitry by CsrA. *J Bacteriol*, 194(1):79–89, 2012.

[357] C. S. Baker, L. A. Eory, H. Yakhnin, J. Mercante, T. Romeo, and P. Babitzke. CsrA inhibits translation initiation of *Escherichia coli* hfq by binding to a single site overlapping the Shine-Dalgarno sequence. *J Bacteriol*, 189(15):5472–81, 2007.

[358] L. C. Martinez, H. Yakhnin, M. I. Camacho, D. Georgellis, P. Babitzke, J. L. Puente, and V. H. Bustamante. Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the *Salmonella* SPI-1 and SPI-2 virulence regulons through HilD. *Mol Microbiol*, 80(6):1637–56, 2011.

[359] H. Yakhnin, A. V. Yakhnin, C. S. Baker, E. Sineva, I. Berezin, T. Romeo, and P. Babitzke. Complex regulation of the global regulatory gene csrA: CsrA-mediated translational repression, transcription from five promoters by Esigma(7)(0) and Esigma(S), and indirect transcriptional activation by CsrA. *Mol Microbiol*, 81(3):689–704, 2011.

[360] X. Wang, A. K. Dubey, K. Suzuki, C. S. Baker, P. Babitzke, and T. Romeo. CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. *Mol Microbiol*, 56(6):1648–63, 2005.

[361] B. L. Wei, A. M. Brun-Zinkernagel, J. W. Simecka, B. M. Pruss, P. Babitzke, and T. Romeo. Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of *Escherichia coli*. *Mol Microbiol*, 40(1):245–56, 2001.

[362] A. V. Yakhnin, C. S. Baker, C. A. Vakulskas, H. Yakhnin, I. Berezin, T. Romeo, and P. Babitzke. CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. *Mol Microbiol*, 87(4):851–66, 2013.

[363] L. M. Patterson-Fortin, C. A. Vakulskas, H. Yakhnin, P. Babitzke, and T. Romeo. Dual posttranscriptional regulation via a cofactor-responsive mRNA leader. *J Mol Biol*, 425(19):3662–77, 2013.

[364] T. Esquerre, M. Bouvier, C. Turlan, A. J. Carpousis, L. Girbal, and M. Cocaign-

Bousquet. The Csr system regulates genome-wide mRNA stability and transcription and thus gene expression in *Escherichia coli*. *Sci Rep*, 6:25057, 2016.

[365] F. M. Barnard, M. F. Loughlin, H. P. Fainberg, M. P. Messenger, D. W. Ussery, P. Williams, and P. J. Jenks. Global regulation of virulence and the stress response by CsrA in the highly adapted human gastric pathogen *Helicobacter pylori*. *Mol Microbiol*, 51(1):15–32, 2004.

[366] S. Bhatt, A. N. Edwards, H. T. T. Nguyen, D. Merlin, T. Romeo, and D. Kalman. The RNA binding protein CsrA is a pleiotropic regulator of the locus of enterocyte effacement pathogenicity island of enteropathogenic *Escherichia coli*. *Infect Immun*, 77(9):3552–68, 2009.

[367] H. Yakhnin, P. Pandit, T. J. Petty, C. S. Baker, T. Romeo, and P. Babitzke. CsrA of *Bacillus subtilis* regulates translation initiation of the gene encoding the flagellin protein (hag) by blocking ribosome binding. *Mol Microbiol*, 64(6):1605–20, 2007.

[368] G. Dugar, S. L. Svensson, T. Bischler, S. Waldchen, R. Reinhardt, M. Sauer, and C. M. Sharma. The CsrA-FliW network controls polar localization of the dual-function flagellin mRNA in *Campylobacter jejuni*. *Nat Commun*, 7:11667, 2016.

[369] D. H. Lenz, M. B. Miller, J. Zhu, R. V. Kulkarni, and B. L. Bassler. CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*. *Mol Microbiol*, 58(4):1186–202, 2005.

[370] T. Weilbacher, K. Suzuki, A. K. Dubey, X. Wang, S. Gudapaty, I. Morozov, C. S. Baker, D. Georgellis, P. Babitzke, and T. Romeo. A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol Microbiol*, 48(3):657–70, 2003.

[371] K. Suzuki, P. Babitzke, S. R. Kushner, and T. Romeo. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. *Genes Dev*, 20(18):2605–17, 2006.

[372] A. Pannuri, C. A. Vakulskas, T. Zere, L. C. McGibbon, A. N. Edwards, D. Georgellis, P. Babitzke, and T. Romeo. Circuitry linking the catabolite repression and Csr global regulatory systems of *Escherichia coli*. *J Bacteriol*, 2016.

[373] C. A. Vakulskas, A. H. Potts, P. Babitzke, B. M. M. Ahmer, and T. Romeo. Regulation of bacterial virulence by Csr (Rsm) systems. *Microbiol Mol Biol Rev*, 79(2):193–224, 2015.

[374] C. K. Maurer, M. Fruth, M. Empting, O. Avrutina, J. Hossmann, S. Nadmid, J. Gorges, J. Herrmann, U. Kazmaier, P. Dersch, R. Muller, and R. W. Hartmann. Discovery of the first small-molecule CsrA-RNA interaction inhibitors

using biophysical screening technologies. *Future Med Chem*, 8(9):931–47, 2016.

[375] F. H. Crick. Codon–anticodon pairing: the wobble hypothesis. *J Mol Biol*, 19(2):548–55, 1966.

[376] A. P. Gerber and W. Keller. An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science*, 286(5442):1146–9, 1999.

[377] F. V. t. Murphy and V. Ramakrishnan. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol*, 11(12):1251–2, 2004.

[378] S. Sheikh, R. Backofen, and Y. Ponty. Impact of the energy model on the complexity of RNA folding with pseudoknots. In *Combinatorial Pattern Matching - 23rd Annual Symposium, CPM 2012, Helsinki, Finland, July 3-5, 2012. Proceedings*, pages 321–333, 2012.

[379] M. G. Seetin and D. H. Mathews. RNA structure prediction: an overview of methods. *Methods Mol Biol*, 905:99–122, 2012.

[380] R. Nussinov, G. Pieczenik, J. R. Griggs, and D. J. Kleitman. Algorithms for loop matchings. *SIAM J Appl Math*, 35(1):68–82, 1978.

[381] T. Xia, J. J. SantaLucia, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37(42):14719–35, 1998.

[382] D. Mathews, J. Sabina, M. Zuker, and D. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, 288(5):911–40, 1999.

[383] D. H. Turner and D. H. Mathews. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res*, 38(Database issue):D280–2, 2010.

[384] M. Zuker and P. Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*, 9(1):133–48, 1981.

[385] R. Lorenz, S. H. Bernhart, C. Höner Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker. ViennaRNA Package 2.0. *Algorithms Mol Biol*, 6:26, 2011.

[386] M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 31(13):3406–15, 2003.

[387] N. R. Markham and M. Zuker. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol*, 453:3–31, 2008.

[388] S. H. Kim, G. J. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J.

Kim, J. Weinzierl, and A. Rich. Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science*, 179(4070):285–8, 1973.

[389] J. E. Ladner, A. Jack, J. D. Robertus, R. S. Brown, D. Rhodes, B. F. Clark, and A. Klug. Structure of yeast phenylalanine transfer RNA at 2.5 A resolution. *Proc Natl Acad Sci USA*, 72(11):4414–8, 1975.

[390] J. Kruger and M. Rehmsmeier. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res*, 34(Web Server issue):W451–4, 2006.

[391] D. Herschlag. RNA chaperones and the RNA folding problem. *J Biol Chem*, 270(36):20871–4, 1995.

[392] L. Rajkowitsch, D. Chen, S. Stampfl, K. Semrad, C. Waldsich, O. Mayer, M. F. Jantsch, R. Konrat, U. Blasi, and R. Schroeder. RNA chaperones, RNA annealers and RNA helicases. *RNA Biol*, 4(3):118–30, 2007.

[393] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–10, 1990.

[394] W. Gerlach and R. Giegerich. GUUGle: a utility for fast exact matching under RNA complementary rules including G-U base pairing. *Bioinformatics*, 22(6):762–4, 2006.

[395] S. Karlin and S. F. Altschul. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc Natl Acad Sci USA*, 87(6):2264–8, 1990.

[396] B. Tjaden, S. S. Goodwin, J. A. Opdyke, M. Guillier, D. X. Fu, S. Gottesman, and G. Storz. Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res*, 34(9):2791–802, 2006.

[397] T. F. Smith and M. S. Waterman. Identification of common molecular subsequences. *J Mol Biol*, 147(1):195–7, 1981.

[398] M. Rehmsmeier, P. Steffen, M. Höchsmann, and R. Giegerich. Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(10):1507–17, 2004.

[399] H. Gong, G.-P. Vu, Y. Bai, E. Chan, R. Wu, E. Yang, F. Liu, and S. Lu. A *Salmonella* small non-coding RNA facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. *PLoS Pathog*, 7(9):e1002120, 2011.

[400] H. Tafer and I. L. Hofacker. RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics*, 24(22):2657–63, 2008.

[401] D. A. Los and N. Murata. Membrane fluidity and its roles in the perception of environmental signals. *Biochim Biophys Acta*, 1666(1-2):142–57, 2004.

[402] I. Konig, A. Zarrine-Afsar, M. Aznauryan, A. Soranno, B. Wunderlich, F. Dingfelder,

J. C. Stuber, A. Pluckthun, D. Nettels, and B. Schuler. Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat Methods*, 12(8):773–9, 2015.

[403] K. Kashefi and D. R. Lovley. Extending the upper temperature limit for life. *Science*, 301(5635):934, 2003.

[404] R. Backofen. Computational Prediction of RNA-RNA Interactions. *Methods Mol Biol*, 1097:417–35, 2014.

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

[406] R. M. Dirks, J. S. Bois, J. M. Schaeffer, E. Winfree, and N. A. Pierce. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Review*, 49(1):65–88, 2007.

[407] M. Andronescu, Z. C. Zhang, and A. Condon. Secondary structure prediction of interacting RNA molecules. *J Mol Biol*, 345(5):987–1001, 2005.

[408] J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29(6-7):1105–19, 1990.

[409] S. Altuvia, A. Zhang, L. Argaman, A. Tiwari, and G. Storz. The *Escherichia coli* OxyS regulatory RNA represses fhlA translation by blocking ribosome binding. *EMBO J*, 17(20):6069–75, 1998.

[410] L. Argaman and S. Altuvia. *fhlA* repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. *J Mol Biol*, 300(5):1101–12, 2000.

[411] S. Boisset, T. Geissmann, E. Huntzinger, P. Fechter, N. Bendridi, M. Possedko, C. Chevalier, A. C. Helfer, Y. Benito, A. Jacquier, C. Gaspin, F. Vandenesch, and P. Romby. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev*, 21(11):1353–66, 2007.

[412] C. Chevalier, S. Boisset, C. Romilly, B. Masquida, P. Fechter, T. Geissmann, F. Vandenesch, and P. Romby. *Staphylococcus aureus* RNAIII binds to two distant regions of *coa* mRNA to arrest translation and promote mRNA degradation. *PLoS Pathog*, 6(3):e1000809, 2010.

[413] U. Mückstein, H. Tafer, J. Hackermüller, S. H. Bernhart, P. F. Stadler, and I. L. Hofacker. Thermodynamics of RNA-RNA binding. *Bioinformatics*, 22(10):1177–82, 2006.

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

[415] D. D. Pervouchine. IRIS: intermolecular RNA interaction search. *Genome Inform*, 15(2):92–101, 2004.

[416] H. Chitsaz, R. Backofen, and S. C. Sahinalp. biRNA: Fast RNA-RNA binding sites prediction. In S. Salzberg and T. Warnow, editors, *Proc. of the 9th Workshop on Algorithms in Bioinformatics (WABI)*, volume 5724 of *Lecture Notes in Computer Science*, pages 25–36. Springer Berlin / Heidelberg, 2009.

[417] A. Pain, A. Ott, H. Amine, T. Rochat, P. Bouloc, and D. Gautheret. An assessment of bacterial small RNA target prediction programs. *RNA Biol*, 12(5):509–13, 2015.

[418] R. Guigo, E. T. Dermitzakis, P. Agarwal, C. P. Ponting, G. Parra, A. Reymond, J. F. Abril, E. Keibler, R. Lyle, C. Ucla, S. E. Antonarakis, and M. R. Brent. Comparison of mouse and human genomes followed by experimental verification yields an estimated 1,019 additional genes. *Proc Natl Acad Sci USA*, 100(3):1140–5, 2003.

[419] S. Washietl and I. L. Hofacker. Identifying structural noncoding RNAs using RNAz. *Curr Protoc Bioinformatics*, Chapter 12:Unit 12.7, 2007.

[420] A. R. Gruber, S. Findeiss, S. Washietl, I. L. Hofacker, and P. F. Stadler. RNAZ 2.0: Improved noncoding rna detection. In *PSB10*, volume 15, pages 69–79, 2010.

[421] S. Washietl, S. Findeiss, S. A. Muller, S. Kalkhof, M. von Bergen, I. L. Hofacker, P. F. Stadler, and N. Goldman. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA*, 17(4):578–94, 2011.

[422] A. X. Li, M. Marz, J. Qin, and C. M. Reidys. RNA-RNA interaction prediction based on multiple sequence alignments. *Bioinformatics*, 27(4):456–63, 2011.

[423] S. E. Seemann, A. S. Richter, T. Gesell, R. Backofen, and J. Gorodkin. PETcofold: predicting conserved interactions and structures of two multiple alignments of RNA sequences. *Bioinformatics*, 27(2):211–219, 2011.

[424] E. Duchaud, C. Rusniok, L. Frangeul, C. Buchrieser, A. Givaudan, S. Taourit, S. Bocs, C. Boursaux-Eude, M. Chandler, J.-F. Charles, E. Dassa, R. Derose, S. Derzelle, G. Freyssinet, S. Gaudriault, C. Medigue, A. Lanois, K. Powell, P. Siguier, R. Vincent, V. Wingate, M. Zouine, P. Glaser, N. Boemare, A. Danchin, and F. Kunst. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat Biotechnol*, 21(11):1307–13, 2003.

[425] H. C. Tekedar, A. Karsi, M. L. Williams, S. Vamenta, M. M. Banes, M. Duke, B. E. Scheffler, and M. L. Lawrence. Complete Genome Sequence of Channel Catfish Gastrointestinal Septicemia Isolate *Edward-*

siella tarda C07-087. *Genome Announc*, 1(6), 2013.

[426] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[427] R. A. Fisher. *Statistical methods for research workers*. Genesis Publishing Pvt Ltd, 1925.

[428] J. Hartung. A note on combining dependent tests of significance. *Biometrical Journal*, 41(7):849–855, 1999.

[429] C. Guan, C. Ye, X. Yang, and J. Gao. A review of current large-scale mouse knockout efforts. *Genesis*, 48(2):73–85, 2010.

[430] G. Prelich. Gene overexpression: uses, mechanisms, and interpretation. *Genetics*, 190(3):841–54, 2012.

[431] C. Soneson and M. Delorenzi. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14:91, 2013.

[432] F. Rapaport, R. Khanin, Y. Liang, M. Pirun, A. Krek, P. Zumbo, C. E. Mason, N. D. Socci, and D. Betel. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol*, 14(9):R95, 2013.

[433] H. Li and R. Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–60, 2009.

[434] S. Hoffmann, C. Otto, S. Kurtz, C. M. Sharma, P. Khaitovich, J. Vogel, P. F. Stadler, and J. Hackermuller. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol*, 5(9):e1000502, 2009.

[435] B. Langmead and S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4):357–9, 2012.

[436] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.

[437] M. Lawrence, W. Huber, H. Pages, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey. Software for computing and annotating genomic ranges. *PLoS Comput Biol*, 9(8):e1003118, 2013.

[438] S. Anders, P. T. Pyl, and W. Huber. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–9, 2015.

[439] Y. Liao, G. K. Smyth, and W. Shi. featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–30, 2014.

[440] M.-A. Dillies, A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin,

N. Servant, C. Keime, G. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, L. Jouneau, D. Laloe, C. Le Gall, B. Schaeffer, S. Le Crom, M. Guedj, and F. Jaffrezic. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform*, 14(6):671–83, 2013.

[441] C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*, 28(5):511–5, 2010.

[442] S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome Biol*, 11(10):R106, 2010.

[443] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12):550, 2014.

[444] M. D. Robinson and A. Oshlack. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, 11(3):R25, 2010.

[445] J. C. Marioni, C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*, 18(9):1509–17, 2008.

[446] J. H. Malone and B. Oliver. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol*, 9:34, 2011.

[447] M. A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G. M. Rubin, J. A. Blake, C. Bult, M. Dolan, H. Drabkin, J. T. Eppig, D. P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J. M. Cherry, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, R. S. Nash, A. Sethuraman, C. L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S. Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E. M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, and R. White. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*, 32 Database issue:D258–61, 2004.

[448] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, 27(1):29–34, 1999.

[449] D. W. Huang, B. T. Sherman, and R. A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1):1–13, 2009.

[450] V. K. Mootha, C. M. Lindgren, K.-F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstrale, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, and L. C. Groop. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*, 34(3):267–73, 2003.

[451] X. Jiao, B. T. Sherman, D. W. Huang, R. Stephens, M. W. Baseler, H. C. Lane, and R. A. Lempicki. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, 28(13):1805–6, 2012.

[452] D. W. Huang, B. T. Sherman, and R. A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1):44–57, 2009.

[453] C. K. Stover, X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, 406(6799):959–64, 2000.

[454] D. Martin, C. Brun, E. Remy, P. Mouren, D. Thieffry, and B. Jacq. GOToolBox: functional analysis of gene datasets based on Gene Ontology. *Genome Biol*, 5(12):R101, 2004.

[455] D. W. Huang, B. T. Sherman, Q. Tan, J. R. Collins, W. G. Alvord, J. Roayaei, R. Stephens, M. W. Baseler, H. C. Lane, and R. A. Lempicki. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*, 8(9):R183, 2007.

[456] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*, 102(43):15545–50, 2005.

[457] A. S. Richter and R. Backofen. Accessibility and conservation: General features of bacterial small RNA-mRNA interactions? *RNA Biol*, 9(7):954–65, 2012.

[458] F. Eggenhofer, H. Tafer, P. F. Stadler, and I. L. Hofacker. RNApredator: fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res*, 39(Web Server issue):W149–54, 2011.

[459] S. Durand, F. Braun, E. Lioliou, C. Romilly, A.-C. Helfer, L. Kuhn, N. Quittot, P. Nicolas, P. Romby, and C. Condon. A nitric oxide regulated small RNA controls expression of genes involved in redox homeostasis in *Bacillus subtilis*. *PLoS Genet*, 11(2):e1004957, 2015.

[460] S. Khandige, T. Kronborg, B. E. Uhlin, and J. Moller-Jensen. sRNA-Mediated Regulation of P-Fimbriae Phase Variation in Uropathogenic *Escherichia coli*. *PLoS Pathog*, 11(8):e1005109, 2015.

[461] K. Baumgardt, K. Smidova, H. Rahn, G. Lochnit, M. Robledo, and E. Evguenieva-Hackenberg. The stress-related, rhizobial small RNA RcsR1 destabilizes the autoinducer synthase encoding mRNA sinI in *Sinorhizobium meliloti*. *RNA Biol*, 13(5):486–99, 2016.

[462] P. Lu, Y. Wang, Y. Zhang, Y. Hu, K. M. Thompson, and S. Chen. RpoS-dependent sRNA RgsA regulates Fis and AcpP in *Pseudomonas aeruginosa*. *Mol Microbiol*, 2016.

[463] D. Balasubramanian, P. T. Ragunathan, J. Fei, and C. K. Vanderpool. A prophage-encoded small RNA controls metabolism and cell division in *Escherichia coli*. *mSystems*, 1(1), 2016.

[464] J. Chen and S. Gottesman. Spot 42 sRNA regulates arabinose-inducible araBAD promoter activity by repressing synthesis of the high-affinity low-capacity arabinose transporter. *J Bacteriol*, 2016.

[465] L. Attaiech, A. Boughammoura, C. Brochier-Armanet, O. Allatif, F. Peillard-Fiorente, R. A. Edwards, A. R. Omar, A. M. MacMillan, M. Glover, and X. Charpentier. Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proc Natl Acad Sci USA*, 2016.

[466] A. Smirnov, K. U. Forstner, E. Holmqvist, A. Otto, R. Gunster, D. Becher, R. Reinhardt, and J. Vogel. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci USA*, 2016.

[467] A. Helwak, G. Kudla, T. Dudnakova, and D. Tollervey. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 153(3):654–65, 2013.

[468] E. Sharma, T. Sterne-Weiler, D. O'Hanlon, and B. J. Blencowe. Global Mapping of Human RNA-RNA Interactions. *Mol Cell*, 62(4):618–26, 2016.

[469] S. A. Waters, S. P. McAteer, G. Kudla, I. Pang, N. P. Deshpande, T. G. Amos, K. W. Leong, M. R. Wilkins, R. Strugnell, D. L. Gally, D. Tollervey, and J. J. Tree. Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *EMBO J*, 2016.

10 Acknowledgments

I dedicate this thesis to Freiburg im Breisgau. To me Freiburg is a town of beauty, joy, knowledge, a home for the last ten years and most importantly the place where I met my best friends Jonas Krisch, Raphael Winkler and Dr. Matthias Kopf.

I am extremely grateful to my principle doctoral advisor Prof. Dr. Rolf Backofen for giving me the opportunity to undertake this course of study under his supervision. Working in his group not only allowed me to be involved in many interesting and challenging scientific projects but also allowed me to develop as a young scientist, thank you Rolf.

I would like to thank my family and especially my parents Dr. Gabriele M. König and Dr. Anthony D. Wright for their support and continued help. I am happy to be able to say that you are not only my parents but also two of my best friends.

Special thanks are owed to Dr. Jens Georg for his invaluable collaborations on all scientific matters as well as for his willingness to share his ideas with me.

Furthermore, I would like to acknowledge Prof. Dr. Rolf Backofen, Prof. Dr. Wolfgang R. Hess, Prof. Dr. Franz Narberhaus, Prof. Dr. Anke Becker, Prof. Dr. Jörg Vogel, Prof. Dr. Kai Papenfort, Dr. Aaron Overlöper, Dr. Marta Robledo, Dr. Erik Holmqvist, Dr. Andreas S. Richter, Dr. Martin Mann, Steffen C. Lott and Dr. Jens Georg for the successful scientific collaborations that led to the completion of this thesis.

I thank Dr. Stephan Klähn, Prof. Dr. Kai Papenfort, Dr. Marta Robledo and Jessica Borgmann for supplying me with the photographs shown in the introductory part of this thesis.

General thanks go to AG Backofen, and especially to Dr. Robert Kleinkauf, Dr. Martin Mann, Dr. Torsten Houwaart and Pavankumar Videm. I hope to always have people like you around.

This list would not be complete without mentioning our hard working secretary Monika Degen-Hellmuth. I think you are the best secretary anyone could ever wish for.