Discovery of RNA/protein complexes by Grad-seq

Ermittlung von RNA/Protein-Komplexen mittels Grad-seq

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Infection and Immunity

submitted by

Jens Hör

from Stuttgart

Würzburg 2020

Submitted on:

Members of the Thesis Committee

Chairperson: Prof. Dr. Thomas Rudel

Primary Supervisor: Prof. Dr. Jörg Vogel

Supervisor (Second): Prof. Dr. Cynthia M. Sharma

Supervisor (Third): Prof. Dr. Utz Fischer

Date of Public Defense:

Date of Receipt of Certificates:

Summary

Complex formation between macromolecules constitutes the foundation of most cellular processes. Most known complexes are made up of two or more proteins interacting in order to build a functional entity and therefore enabling activities which the single proteins could otherwise not fulfill. With the increasing knowledge about noncoding RNAs (ncRNAs) it has become evident that, similar to proteins, many of them also need to form a complex to be functional. This functionalization is usually executed by specific or global RNA-binding proteins (RBPs) that are specialized binders of a certain class of ncRNAs. For instance, the enterobacterial global RBPs Hfq and ProQ together bind >80 % of the known small regulatory RNAs (sRNAs), a class of ncRNAs involved in post-transcriptional regulation of gene expression.

However, identification of RNA-protein interactions so far was performed individually by employing low-throughput biochemical methods and thereby hindered the discovery of such interactions, especially in less studied organisms such as Gram-positive bacteria. Using gradient profiling by sequencing (Grad-seq), the present thesis aimed to establish high-throughput, global RNA/protein complexome resources for *Escherichia coli* and *Streptococcus pneumoniae* in order to provide a new way to investigate RNA-protein as well as protein-protein interactions in these two important model organisms.

In *E. coli*, Grad-seq revealed the sedimentation profiles of 4,095 (~85 % of total) transcripts and 2,145 (~49 % of total) proteins and with that reproduced its major ribonucleoprotein particles. Detailed analysis of the in-gradient distribution of the RNA and protein content uncovered two functionally unknown molecules—the ncRNA RyeG and the small protein YggL—to be ribosome-associated. Characterization of RyeG revealed it to encode for a 48 aa long, toxic

protein that drastically increases lag times when overexpressed. YggL was shown to be bound by the 50S subunit of the 70S ribosome, possibly indicating involvement of YggL in ribosome biogenesis or translation of specific mRNAs.

S. pneumoniae Grad-seq detected 2,240 (~88 % of total) transcripts and 1,301 (~62 % of total) proteins, whose gradient migration patterns were successfully reconstructed, and thereby represents the first RNA/protein complexome resource of a Gram-positive organism. The dataset readily verified many conserved major complexes for the first time in *S. pneumoniae* and led to the discovery of a specific interaction between the 3'→5' exonuclease Cbf1 and the competence-regulating cia-dependent sRNAs (csRNAs). Unexpectedly, trimming of the csRNAs by Cbf1 stabilized the former, thereby promoting their inhibitory function. *cbf1* was further shown to be part of the late competence genes and as such to act as a negative regulator of competence.

Zusammenfassung

Makromoleküle, die Komplexe bilden, sind die Grundlage der meisten zellulären Prozesse. Die meisten bekannten Komplexe bestehen aus zwei oder mehr Proteinen, die interagieren, um eine funktionelle Einheit zu bilden. Diese Interaktionen ermöglichen Funktionen, die die einzelnen Proteine nicht erfüllen könnten. Wachsende wissenschaftliche Erkenntnisse über nichtkodierende RNAs (ncRNAs) haben gezeigt, dass, analog zu Proteinen, auch viele ncRNAs Komplexe bilden müssen, um ihre Funktionen ausüben zu können. Diese Funktionalisierung wird normalerweise von spezifischen oder globalen RNA-bindenden Proteinen (RBPs), die auf eine bestimmte Klasse an ncRNAs spezialisiert sind, durchgeführt. So binden beispielsweise die in Enterobakterien verbreiteten globalen RBPs Hfq und ProQ zusammen >80 % der bekannten kleinen regulatorischen RNAs (sRNAs)—eine Klasse der ncRNAs, die in die posttranskriptionelle Genexpressionsregulation involviert ist.

RNA-Protein-Interaktionen wurden bisher anhand einzelner Moleküle und mithilfe von biochemischen Methoden mit niedrigem Durchsatz identifiziert, was die Entdeckung solcher Interaktionen erschwert hat. Dies gilt insbesondere für Organismen, die seltener Gegenstand der Forschung sind, wie beispielsweise grampositive Bakterien. Das Ziel dieser Doktorarbeit war es, mittels *gradient profiling by sequencing* (Grad-seq) globale Hochdurchsatzkomplexomdatensätze der RNA-Protein-Interaktionen in *Escherichia coli* und *Streptococcus pneumoniae* zu generieren. Diese Datensätze ermöglichen es auf eine neue Art und Weise RNA-Protein- und Protein-Protein-Interaktionen in diesen wichtigen Modellorganismen zu untersuchen.

Die *E. coli* Grad-seq-Daten beinhalten die Sedimentationsprofile von 4095 Transkripten (∼85 % des Transkriptoms) und 2145 Proteinen (∼49 % des Proteoms), mit denen die wichtigsten Ribonukleoproteine reproduziert werden konnten. Die

detaillierte Analyse der Verteilung von RNAs und Proteinen im Gradienten zeigte, dass zwei Moleküle, deren Funktionen bisher unbekannt waren—die ncRNA RyeG und das kleine Protein YggL—ribosomenassoziiert sind. Durch weitere Charakterisierung konnte gezeigt werden, dass RyeG für ein toxisches Protein mit einer Länge von 48 Aminosäuren kodiert, das bei Überexpression die Latenzphase drastisch verlängert. Für YggL konnte eine Interaktion mit der 50S Untereinheit von 70S Ribosomen nachgewiesen werden, was auf eine potenzielle Funktion in der Biogenese von Ribosomen oder bei der Translation bestimmter mRNAs hindeutet.

Die *S. pneumoniae* Grad-seq Daten beinhalten 2240 Transkripte ($\sim$88 % des Transkriptoms) und 1301 Proteine ($\sim$62 % des Proteoms), deren Migrationsprofile im Gradienten erfolgreich rekonstruiert werden konnten. Dieser RNA/Protein-Komplexomdatensatz eines grampositiven Organismus ermöglichte erstmalig die Verifizierung der wichtigsten konservierten Komplexe von *S. pneumoniae*. Weiterhin konnte eine spezifische Interaktion der 3′→5′-Exonuklease Cbf1 mit den cia-dependent sRNAs (csRNAs), die an der Regulation von Kompetenz beteiligt sind, nachgewiesen werden. Überraschenderweise stabilisiert das von Cbf1 durchgeführte Kürzen der csRNAs die selbigen, was deren inhibitorische Funktion unterstützt. Darüber hinaus konnte gezeigt werden, dass *cbf1* eines der späten Kompetenzgene ist und als solches als negativer Regulator der Kompetenz agiert.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Bacterial complexes

Functional interactions between macromolecules build the foundation of life. By interacting, proteins and RNA are able to carry out functions they would not be able to do individually. Catalytic pockets are formed, performance of successive reactions is increased by bringing catalytic centers in proximity, membrane pores are formed and protein activities are modulated, just to name a few functions arising from complex formation. The overall result of this impressive feat of evolution is that complexes carry out the most fundamental processes in a cell: In *Escherichia coli*, DNA is replicated by the DNA polymerase III holoenzyme, a complex containing 10 different subunits and 17 proteins in total (Fijalkowska et al., 2012). Transcription of DNA is then performed by RNA polymerase (RNAP) containing 4 core subunits and whose specificity gets modulated by 7 different σ-factors and the noncoding 6S RNA (Feklístov et al., 2014; Sutherland and Murakami, 2018; Wassarman, 2018). Following transcription, translation is carried out by the ribosome, the most sophisticated complex in a bacterial cell, which is a ribozyme built up by >50 different proteins and 3 noncoding RNAs (ncRNAs) (Davis and Williamson, 2017). Not mentioning many other essential metabolic pathways in the cell, these three examples alone show the importance of macromolecules forming higher order interactions to fulfill their functions.

Apart from protein-protein complexes, which have been studied since the dawn of biochemistry, bacterial RNA/protein complexes, or ribonucleoprotein particles (RNPs), have gained much more attention during the last decades. This is in part due to the innovation of new methodologies with which to study these interactions and in part due to the appreciation of RNA being more than just the carrier of genetic information. Classic RNPs such as the aforementioned ribosome, the signal recognition particle (Akopian et al., 2013) or RNase P (Mondragón, 2013) all involve RNA-binding proteins (RBPs) that specifically bind their target RNAs with specialized RNA-binding domains. Similar to protein-protein complexes, the formation of RNPs allows functionalization of the otherwise inactive subunits. Another important use bacteria make of RNPs is to rapidly control their gene expression in response to environmental changes, allowing adaptation and survival (Babitzke et al., 2019; Holmqvist and Vogel, 2018). This process often involves an RBP that forms a complex with both a small regulatory RNA (sRNA) and its target mRNA in order to provide a platform for the regulation to happen (Hör et al., 2020b). Thus, RNPs are of similar importance as protein-protein complexes with which together they form the whole of the bacterial complexome. Understanding and identification of both the players within the complexome and the complexome as a whole requires sophisticated methods to study these intricate cellular machineries.

1.2   Investigation of macromolecular interactions

The analysis and identification of protein-protein interactions (PPIs) and RNPs poses several challenges. When one or more subunits of a complex are known, many biochemical methods exist that allow the identification of additional subunits. These methods usually rely on purification of the known subunit under conditions that allow to wash away non-interacting molecules. Subsequent analysis of the bound fraction then enables identification of the whole complex. However, *de novo* identification of complexes is more complicated since none of the subunits are known. This is especially true for understudied organisms that are distant to the most studied model organisms and are therefore lacking homologs of known complexes. These circumstances make it highly desirable to introduce new methods that are capable

of unbiasedly analyzing the complexome of an organism of interest in order to predict complex formation. The following sections will give an overview of different types of methods that can be applied to investigate different types of complexes in bacteria.

1.2.1 Bait-based methods: protein-protein interactions

As stated before, many methods exist to establish the interactome of a protein of interest. Since specific antibodies are usually not available for bacterial proteins, the bait protein most of the time is tagged in order to allow purification. One classical approach is affinity purification followed by mass spectrometry (AP/MS), for which cells are lysed and the resulting lysates incubated with a matrix allowing immobilization of the bait protein via, *e.g.*, a tag-specific antibody (Kwan and Emili, 2016). After several washing steps removing unbound molecules, the complexes are eluted and the bound fraction analyzed by mass spectrometry (MS) (Figure 1.1 A).

Another widely used method to study PPIs is the yeast two-hybrid (Y2H) system (Koegl and Uetz, 2007; Rajagopala et al., 2012). Y2H takes advantage of a transcription factor (TF) that possesses two domains: a DNA-binding domain (BD) and an activation domain (AD). Usually, the BD binds to an upstream activating sequence (UAS) of a reporter gene whose transcription can then be activated by the AD. To study PPIs, the TF is split into its two domains, thereby rendering it inactive when expressed in yeast. The BD is then fused to a protein of interest (called the bait) and the AD is fused to proteins (called the preys) whose interactions with the bait are to be tested. If a sufficiently strong interaction occurs between the bait and the prey, function of the TF is restored and the reporter gene turned on, indicating a functional PPI (Figure 1.1 B).

1.2.2 Bait-based methods: RNA-protein interactions

The methods mentioned so far exclusively investigate PPIs. For the global investigation of the interaction between RNA and protein, there are two different approaches, which, conceptually, are very similar to AP/MS: pulling down an RBP of interest and sequencing the bound RNA or pulling down an RNA of interest and identifying the bound proteins via MS.
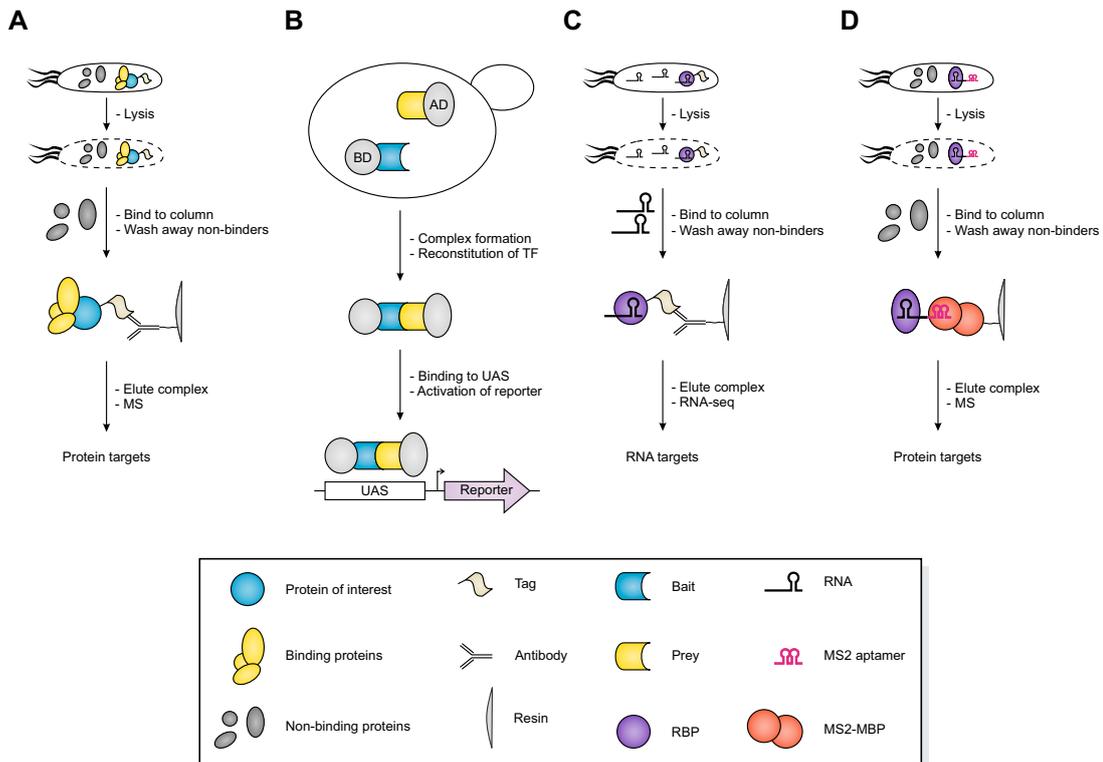
Figure 1.1: Non-global interactome methods. (**A**) Affinity purification followed by mass spectrometry (AP/MS). Following lysis of a culture of cells, the protein of interest (POI) is immobilized on a column, usually by antibody-dependent capture of a tag placed on the POI. Next, proteins that do not bind the POI are washed away and the complex of the POI and its specific binder(s) is eluted. Identification of the binders is performed via MS. (**B**) Yeast two-hybrid (Y2H). Interaction between two proteins is tested by fusing one to a binding domain (BD) and the other to an activating domain (AD). The two fusion proteins are expressed in yeast, allowing the proteins to interact. If they do, the BD and AD form a functional transcription factor (TF) able to bind an upstream activating sequence (UAS), triggering the expression of a reporter that confirms the interaction. (**C**) RNA-immunoprecipitation followed by RNA-seq (RIP-seq). Same procedure as in (**A**), but the POI in this case is an RNA-binding protein (RBP). After elution, the RNA targets of the RBP are identified via RNA-seq. (**D**) MS2 pull-down. An RNA is tagged with the MS2 aptamer and expressed *in vivo*. After lysis of the culture, the MS2-tagged RNA is immobilized by binding to a column-bound MS2-maltose binding protein (MBP)-hybrid protein. Proteins that do not interact with the MS2-tagged RNA are washed away and the remaining complexes are eluted. Identification of protein targets of the MS2-tagged RNA are identified via MS.

1.2.2.1 Protein-centric methods

The most basic method that allows analysis of the interactome of an RBP of interest is RNA immunoprecipitation followed by RNA-seq (RIP-seq) (Hör et al., 2018). To allow affinity purification, the RBP is usually tagged, which is preferentially done on the chromosome to avoid changes in expression, which could otherwise lead to an altered interactome profile. After harvesting the cells at the desired time point or condition, a standard affinity purification is performed and the interacting RNA purified by means of organic phase extraction. The RBP interactome can then be analyzed by sequencing of the bound RNA and comparing it to the untagged wild type to exclude unspecifically bound transcripts (Figure 1.1 C).

One of the limitations of RIP-seq is that it enriches full-length transcripts and therefore gives no additional information about the potential binding site or binding motif of the RBP within the transcripts. To overcome this, crosslinking immunoprecipitation followed by RNA-seq (CLIP-seq, a.k.a. HITS-CLIP) was developed (Andresen and Holmqvist, 2018; Hör et al., 2018; Lee and Ule, 2018). Here, irradiation with UV light of 254 nm wavelength is performed before the cells are harvested, leading to *in vivo* crosslinking of RNA to their binding proteins. This covalent bond allows digestion of RNA that is not protected by the RBP using ribonucleases (RNases) during the subsequent affinity purification. It further allows more stringent washing conditions as compared to RIP-seq, enriching direct targets of the RBP. After purification of the bound RNA fragments, RNA-seq is performed. The recovered RNA sequences, so-called peaks, crucially enable the search for potential binding motifs of the RBP, which is extremely valuable for the understanding of the RBP and which is not possible by RIP-seq.

Furthermore, the crosslinked nucleotide often leads to errors during reverse transcription and therefore allows to estimate which nucleotide within a peak was bound by the RBP. Many different modifications of the CLIP-seq protocol exist (Lee and Ule, 2018), the most important of which are RNA interaction by ligation and sequencing (RIL-seq) and UV-crosslinking, ligation and sequencing of hybrids (CLASH), both of which add a ligation step to the CLIP protocol, allowing the identification of RNA-RNA interactions formed on or by the RBP (Hör and Vogel, 2017; Kudla et al., 2011; Melamed et al., 2018). Furthermore, as an alternative to UV irradiation, formalde-

hyde CLIP-seq (fCLIP-seq) uses formaldehyde to crosslink RNA-protein interactions *in vivo* (Kim et al., 2017; Kim and Kim, 2019). This method improves crosslinking of double-stranded RNA (dsRNA) to dsRNA-binding proteins (Kim et al., 2017; Ricci et al., 2014) but has the disadvantage of also crosslinking protein-protein interactions, which might cause artifacts (Lee and Ule, 2018).

1.2.2.2   RNA-centric methods

If the objective is to identify RBPs rather than the targets of a known RBP, RNA can be used as bait and the bound proteins analyzed by MS. Originally developed to purify eukaryotic RNPs (Bardwell and Wickens, 1990; Bessonov et al., 2008; Zhou et al., 2002), the most common method to perform RNA-centric pull-downs in bacteria is the MS2 approach (Corcoran et al., 2012; Lalaouna et al., 2015; Said et al., 2009; Smirnov et al., 2016). For this, the phage-derived MS2 aptamer is added to the 5' or 3' end of the RNA of interest and either expressed *in vivo* or synthesized *in vitro* and added to a cell lysate. In the next step, the RNA/protein complexes are captured by binding of the aptamer to the MS2 coat protein, which is fused to a maltose-binding protein, allowing immobilization on an amylose resin. After several washes, the bound proteins are recovered, analyzed by MS and the enrichment compared to the control (the MS2 aptamer itself) is calculated (Figure 1.1 D). The power of this approach lies within using many different RNAs, *e.g.*, functionally related ones, for the pull-down in order to discover potential common or even global RBPs.

1.2.3   Global methods: binary methods

While all of the mentioned protein- and RNA-centric methods give invaluable insight into the interactomes of specific molecules of interest, none of them is suitable for the description of the whole complexome of a cell. Therefore, several methods have been developed that analyze the interactions within a cell at a global level.

One option to achieve the global description of the complexome is to scale up binary methods. One of the most popular methods for the global investigation of PPIs is Y2H (Brückner et al., 2009). To achieve this, Y2H is performed as previously described, but instead of only using a limited number of baits and preys, the whole ORFeome (ORF = open reading frame) of the organism of interest is cloned as both
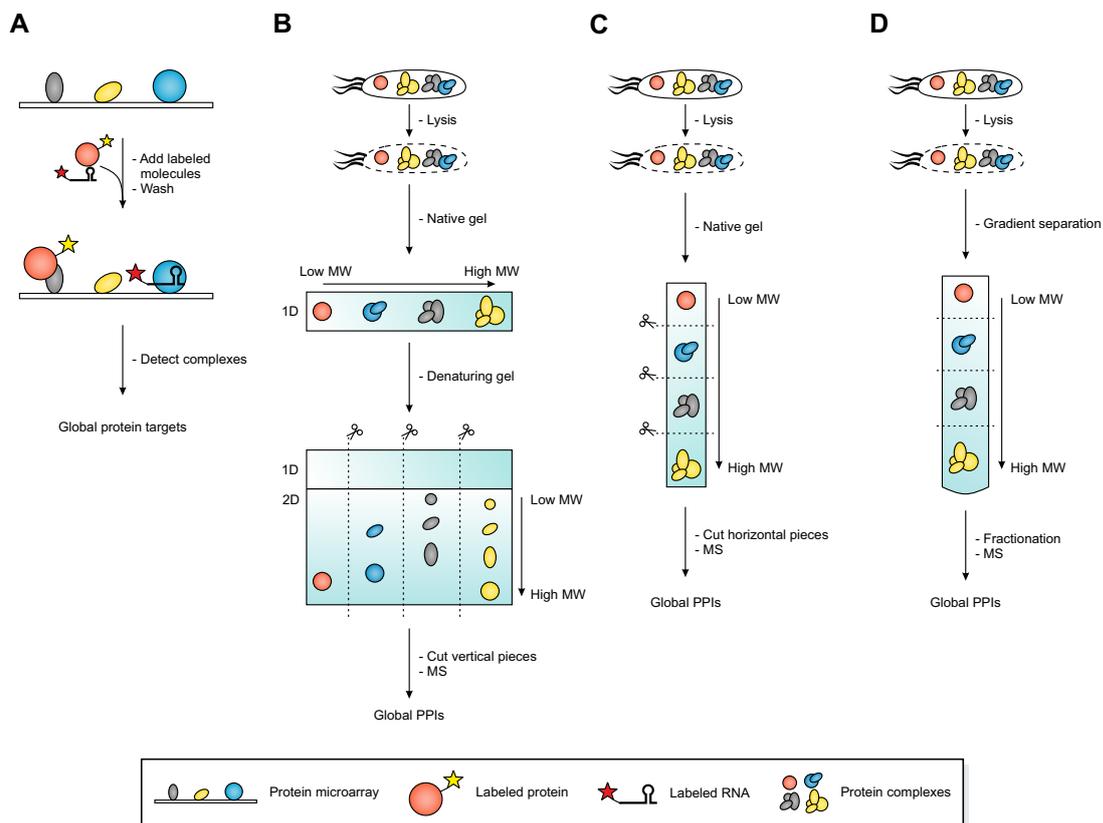
Figure 1.2: Global interactome methods. (**A**) Protein microarray. The proteome of the organism of interest is immobilized on a chip. Labeled proteins and/or RNAs are added to the chip, followed by a washing step and detection of the label. Based on the position of the signal on the chip, the target protein of the labeled molecule can be identified. (**B**) 2D gel analysis followed by mass spectrometry (2D-MS). Following lysis of a culture of cells, the lysate is first run on a native gel, separating complexes and then on a denaturing gel, separating the subunits of the complexes. MS of vertical gel slices identifies the subunit composition of the complexes. (**C**) Complexome profiling. Following lysis of a culture of cells, the lysate is run on a high-resolution native gel, separating complexes by size. The gel is then cut into horizontal pieces and their protein content identified via MS. (**D**) Protein correlation profiling (PCP). Following lysis of a culture of cells, the lysate is run on a gradient, separating the complexes. The gradient is then fractionated and the protein content of each fraction is identified via MS.

bait and prey, allowing an all-against-all analysis of PPIs. Interactions recovered with both ORFs being used as bait and prey are especially meaningful. Still, this method has several drawbacks. Firstly, the number of false-negatives is very high ($\sim$70 % in *E. coli* (Rajagopala et al., 2014)) due to low assay sensitivity. Secondly, false-positives can also be an issue and are usually corrected for by not considering proteins with too many interactions (Rajagopala et al., 2014). Nevertheless, global Y2H provides a great starting point for more detailed analyses of complexes and can be used for functional predictions of interactions.

Similar to global Y2H, AP/MS can also be scaled up to include the whole ORFeome as baits. Different approaches exist, which mostly differ in the way the tagged proteins are expressed: natively or by overexpression. Both options have their advantages and disadvantages; *e.g.,* native expression tends to have less false-positives in exchange for less proteins that can be purified, while overexpression increases the number of baits that can be purified in exchange for more false-positives (Arifuzzaman et al., 2006; Butland et al., 2005). Global AP/MS is very sensitive and generally identifies a considerably larger number of interactions than Y2H, making it absolutely necessary to further refine the predictions by genomic context or downstream validations (Rajagopala et al., 2014).

Further, since AP/MS is able to pull-down complexes binding to the bait, it is not a truly binary method, making it difficult to discriminate between direct and indirect interactions. Then again, this property can lead to the purification and thereby description of whole complexes. In comparison to global Y2H, global AP/MS has the advantage of being performed *in vivo*, which eliminates potential issues with, *e.g.,* protein folding that can arise due to expression in a different species (Berggård et al., 2007). This allows AP/MS to investigate membrane complexes on a global scale (Babu et al., 2018), which generally is a difficult task since membrane proteins often have limited or no solubility.

Protein microarrays, and more specifically functional protein microarrays, are a completely different approach. Here, the interactions are not investigated within a cell but *in vitro* on a solid surface like a glass slide (Sutandy et al., 2013). For this, the whole proteome of the organism of interest is immobilized on a chip, allowing interaction studies by adding labeled protein or RNA. After washing away unbound

molecules, interactions can be detected by scanning the chip for signals of the labeled molecules (Figure 1.2 A). While protein microarrays are not a truly global method since all-against-all interaction studies are difficult to perform, they do allow the *de novo* investigation of RBP-RNA interactions and with that the discovery of RBPs (Sutandy et al., 2016). Yet, they suffer from a similar issue as Y2H: proper folding and function of the purified proteins cannot be guaranteed, increasing the chances for false-negative results. Production and handling of protein microarrays is expensive and difficult, which probably is the reason why the only bacterium they have been used for so far is *E. coli* (Chen et al., 2008; Thao et al., 2010).

1.2.4 Global methods: non-binary methods

To get a better overview of the complexome of a cell, non-binary methods that can globally describe intact complexes, rather than only the interactions between their subunits, have to be used. Apart from the throughput, another major advantage of a global approach is that tagging of proteins/RNAs is not necessary, which should reduce biases that may be introduced by the tag. This kind of analysis is rather simple when investigating a single complex, *e.g.*, to find out more about its subunit stoichiometry, but gets complicated when the whole complexome of a cell is to be investigated.

1.2.4.1 2D gel analysis followed by MS (2D-MS)

One solution to tackle this hurdle is to perform a 2D gel analysis of a cell lysate that first separates the complexes according to the total size on a blue/colorless native gel and then by size of the subunits using denaturing SDS-PAGE (Lasserre et al., 2006; Pan et al., 2010, 2011). The resulting gel contains complexes in the x-dimension and their corresponding subunits in the y-dimension. After identification of a complex via gel staining, its identity can be revealed using MS (Figure 1.2 B). With that, 2D-MS allows a rather fast overview of the complexome and even provides a platform for the analysis of membrane complexes. It is limited, however, in several ways: Staining can be biased by the properties of individual proteins, making it difficult to correctly assign spots on the gel to a complex. Moreover, the resolution in the first dimension is rather low (Lasserre et al., 2006).

1.2.4.2   Complexome profiling

A similar approach to 2D-MS is complexome profiling. Instead of running a 2D gel followed by visual identification of complexes by vertical alignment of spots, complexome profiling omits the second, denaturing dimension and instead runs a longer, higher resolution native gel (Diéguez-Casal et al., 2014; Gorka et al., 2019; Rugen et al., 2019; Senkler et al., 2017; Wöhlbrand et al., 2016). In a second step, the resulting gel is cut into many pieces and each piece is analyzed by MS. This allows the reconstruction of migration profiles of all detectable proteins within the gel (Figure 1.2 C). The advantage over 2D-MS is that the gel is basically "fractionated", achieving much higher resolution. Based on this, predictions can be made about the composition of complexes or the involvement of individual proteins within complexes based on the "guilt-by-association" logic.

Going one step further, cryo-slicing blue native-MS (csBN-MS) uses cryo-microtome slicing in order to cut a frozen blue native gel into >200 equal pieces, which are subsequently analyzed by MS (Müller et al., 2016, 2019). The drastic increase in fraction numbers compared to traditional complexome profiling lets csBN-MS go close to or even reach the resolution limit of a blue native gel. So far, complexome profiling has not been used to investigate RNPs and it seems difficult to purify both proteins and RNA from the small pieces of gel generated by the protocol. It should be possible, however, making complexome profiling an interesting candidate to globally study RNPs in the future.

1.2.4.3   Protein correlation profiling

Similar to complexome profiling, protein correlation profiling (PCP) makes use of the "guilt-by-association" logic that identifies interacting proteins based on co-elution. PCP, however, employs density gradients or size exclusion chromatography (SEC) instead of gels in order to achieve fractionation (Figure 1.2 D). It was first used in eukaryotes to assign proteins to specific organelles (Andersen et al., 2003; Foster et al., 2006), but has been applied to bacteria as well (Carlson et al., 2019). Similar to complexome profiling, PCP does not require any tagging, thereby enabling the use of native lysates. Moreover, the use of gradients or SEC greatly simplifies the down-

stream extraction of macromolecules, which makes PCP a global method that can analyze RNPs, as was shown in a study using a related method (Rederstorff et al., 2010; Rederstorff and Hüttenhofer, 2011). Lastly, the combination of several separation methods can improve the resolution of PCP, though it requires the complexes to be stable enough to survive the process (Crozier et al., 2017; Gazestani et al., 2016; Havugimana et al., 2012).

1.2.5 Gradient profiling by sequencing (Grad-seq)

A conceptually similar method to PCP is gradient profiling by sequencing (Grad-seq) (Smirnov et al., 2016, 2017a), which is the main focus of the present thesis. In Grad-seq, gradient centrifugation is used to separate the content of a wild-type lysate. Subsequently, the gradient is fractionated and the RNA and protein content of each fraction is analyzed by RNA-seq and liquid chromatography-tandem MS (LC-MS/MS), respectively, allowing the reconstruction of how each detectable molecule migrated through the gradient. These data then allow the analysis of the complexome of the used organism, since stable complexes are expected to travel through the gradient as one particle, resulting in congruent migration profiles of their subunits (Figure 1.3).

The following sections will give a brief historic overview of the development of gradient centrifugation as a means to separate the cellular content, the theory behind it and how it is used in Grad-seq.

1.2.5.1 Brief history of ultracentrifugation as a means to purify cellular particles

"The new centrifuge constructed by us allows the determination of particles that cannot be made visible in the ultra-microscope. In analogy with the naming of the ultra-microscope and ultra-filtration apparatus we propose the name ultra-centrifuge for this apparatus." With this quote from Svedberg and Rinde's paper, in which they first described their invention of the ultracentrifuge (Svedberg and Rinde, 1924), much in the field of biochemistry and cell biology should change in the years to come. While the authors originally developed the ultracentrifuge in order to study particles in colloid solutions, it soon became evident that it could also be used for the separation of the cellular content.
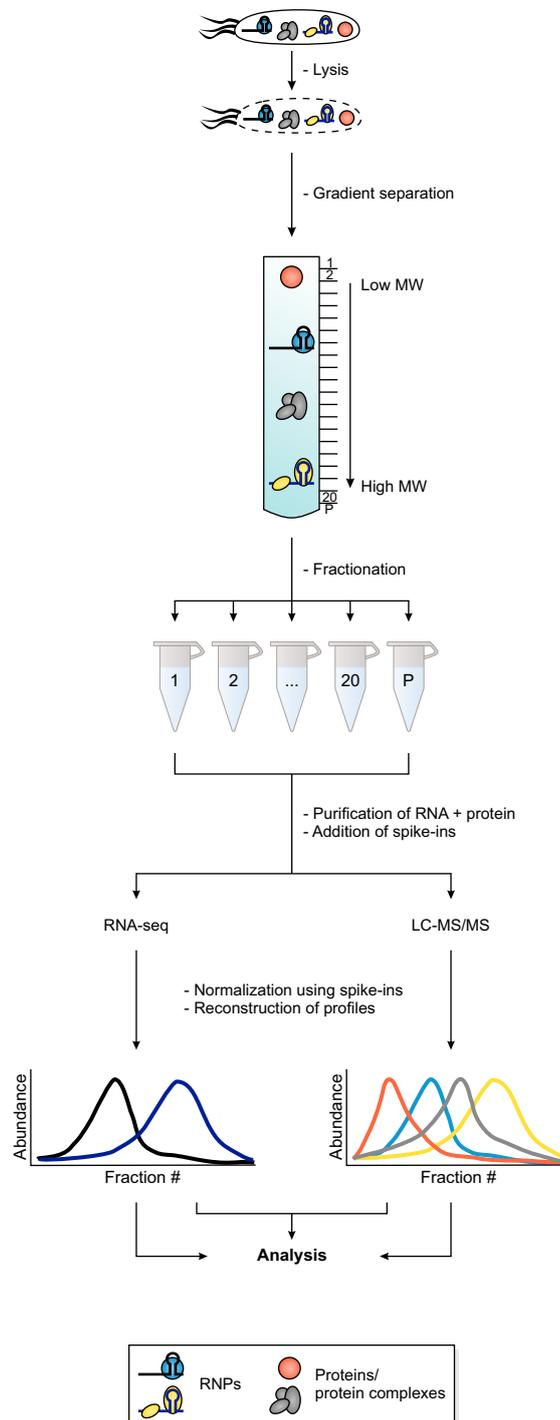
Figure 1.3: Gradient profiling by sequencing (Grad-seq). Following lysis of a culture of cells, the lysate is run on a gradient, separating the complexes. After fractionation of the gradient, the RNA and protein content of each fraction is purified, complex spike-ins are added and RNA-seq and MS are performed for each fraction. Subsequent normalization of the global data allows the reconstruction of the sedimentation profiles of all detectable RNAs and proteins.

The first major breakthrough in cell fractionation using ultracentrifugation was achieved when Albert Claude revolutionized the method of differential centrifugation (Claude, 1946a,b). This was the first time consecutive centrifugation steps were applied to fractionate the cytoplasmic content of cells into three major fractions: "Large granules" (mostly mitochondria), "microsomes" (remnants of the endoplasmic reticulum) and "supernate" (the soluble supernatant). Basically, the protocol consisted of the centrifugation of lysed cells to sediment "large granules" at low centrifugal force, followed by ultracentrifugation of the supernatant at high centrifugal force in order to sediment "microsomes", leaving behind the "supernate".

In this work, the fractions were also characterized according to their specific content. The author mentions, for example, that RNA and RNPs were mostly present in the "large granules" and "microsomes" fractions. By improving differential centrifugation, nuclei were isolated for the first time soon after and the RNA content of the different fractions quantified, revealing that more than 50 % of the total RNA was present in the "microsomes" (Hogeboom et al., 1948; Schneider, 1948). By nowadays' standards, this method seems rather unspectacular. However, owing to its simplicity and affordability, it is still widely used for the crude purification of organelles, which shows the importance and impact of this method, especially considering it was developed more than 70 years ago.

1.2.5.2   Density gradient centrifugation

An important improvement to cell fractionation was introduced with the development of density gradient centrifugation in the early 1950s. Interestingly, this method originally came from the field of plant virology and was first used to purify potato yellow dwarf virus from a virus concentrate created by differential centrifugation (Brakke, 1951). In this important note, the author floated a thin layer of the virus concentrate on a sucrose gradient, which increased in density down the tube, and centrifuged the solution so the particles within the concentrate would sediment as zones within the gradient. Centrifugation was stopped before the particles reached density equilibria, allowing separation according to the sedimentation rate of the particles (more on this below). After centrifugation, a clear zone of virus particles could be detected by means of scattered light and only this zone contained infective

potential when applied to leaves—the other fractions of the gradient were unable to cause lesions.

Numerous modifications of this method were developed soon after it was published. For sake of simplicity, only the two main categories and their applications will be highlighted.

1.2.5.3 Isopycnic gradient centrifugation

In isopycnic gradient centrifugation, particles are separated by centrifugation until they reach their isopycnic positions within a density gradient (Brakke, 1961; Anderson, 1955). This is the position where the opposing processes of sedimentation and diffusion have reached an equilibrium and the sum of forces on the particle is zero (Meselson et al., 1957). The density of the gradient is equal to that of the particle at this position and the separation is solely based on the density of the particle. This also means that the density of the gradient has to be greater than the density of the particles to be separated. The major achievement of this method clearly is the formal proof of the semiconservative DNA replication model. For this, Meselson and Stahl used self-forming CsCl gradients to separate ^{14}N- and ^{15}N-labeled DNA from each other, which allowed them to observe ^{14}N incorporation into DNA after a switch from ^{15}N-containing medium (Meselson and Stahl, 1958).

1.2.5.4 Rate zonal centrifugation

Rate zonal centrifugation is the method of choice for Grad-seq and is also referred to as gradient differential centrifugation (Brakke, 1961; Anderson, 1955). This variant separates particles, which are floated on a preformed gradient, into zones. The gradient fulfills the purpose of stabilizing the forming zones by preventing convection of the particles. Centrifugation is stopped before the particles reach their isopycnic positions, separating them according to their specific sedimentation rates, which are based on their masses, sizes and shapes. The relationship between the sedimentation rate (*i.e.*, the sedimentation velocity) of a particle and its mass, size and shape is given by the particle's sedimentation coefficient, which can be calculated using the Svedberg formula (Svedberg and Pedersen, 1940):

$$S = \frac{v_t}{\omega^2 r}, \tag{1.1}$$

where S is the sedimentation coefficient (unit Svedberg; $1S = 10^{-13}$ s); v_t is the terminal sedimentation velocity of the particle during centrifugation (unit $\frac{m}{s}$) and $\omega^2 r$ is the centrifugal acceleration, where ω is the angular velocity of the rotor (unit $\frac{1}{s}$) and r the distance of the particle to the axis of the rotor (unit m).

The sedimentation rate v_t is defined as the terminal velocity the particle reaches as soon as the centrifugal force is balanced by the friction of the particle within the fluid:

$$v_t = \frac{M(1 - \bar{v}\rho)\omega^2 r}{f}, \tag{1.2}$$

where M is the mass of the particle (unit kg); $\bar{v}$ is the partial specific volume of the particle (*e.g.*, typically $0.73 \frac{cm^3}{g}$ for proteins and $0.55 \frac{cm^3}{g}$ for DNA); ρ is the density of the solvent ($1.0 \frac{g}{cm^3}$ for water at $20\,°C$) and f is the frictional coefficient of the particle (unit $\frac{g}{s}$).

This formula immediately shows that the sedimentation rate is increasing with the effective mass of the particle in the medium. It also shows why a particle at isopycnic position has a sedimentation rate of zero: When the density of the particle (the invers of its partial specific volume) and the density of the medium are equal, the equation equals zero. To understand how the size and the shape of a particle influence sedimentation, the frictional coefficient has to be looked at in more detail:

$$f = \frac{6\pi\eta}{R_s}, \tag{1.3}$$

where η is the viscosity of the solvent ($0.01 \frac{g}{cm^{-s}}$ for water at $20\,°C$) and R_s is the Stokes radius (unit nm).

The frictional coefficient depends on the Stokes radius, which is the radius of a smooth sphere that would give the corresponding value for f for the particle of interest. Since biological molecules like proteins are never perfectly spherical or smooth, but are rather only approximately globular or even elongated in shape and their surface is rather rough, R_s will always be larger than the minimal radius of a sphere that could contain the mass of the particle of interest (Erickson, 2009). To

put it into other words, R_s considers both the asymmetry of the particle and its hydration shell (which increases the particle's effective size). If v_t in formula 1.1 is now replaced by formula 1.2, the resulting formula for the sedimentation coefficient is

$$S = \frac{M(1 - \bar{v}\rho)\omega^2 r}{\omega^2 r f} = \frac{M(1 - \bar{v}\rho)}{f}. \tag{1.4}$$

With this, it is evident that the sedimentation coefficient is only dependent on the mass and the frictional coefficient of the particle in the medium. Therefore, an increase in mass of the particle, as stated before, increases S and an increase in size or asymmetry (*i.e.*, shape) of the particle decreases S. Rearranging formula 1.1 gives the final explanation on the relationship between a particle's sedimentation rate and its properties:

$$v_t = S\omega^2 r, \tag{1.5}$$

which shows that the sedimentation rate is dependent on the sedimentation coefficient and the applied centrifugal force. In practice, this can easily be observed by the fact that the 50S large ribosomal subunit will travel further in a rate zonal centrifugation experiment than the 30S small ribosomal subunit because it will reach a higher velocity. Lastly, it is important to note that the density and viscosity of the solvent is dependent on the temperature and the type of solvent, leading to changes in S values. Therefore, S values in the literature are usually given for water at 20 °C (often depicted as S_w^{20})(Erickson, 2009).

1.2.5.5   Specific considerations for Grad-seq: gradient preparation

To use rate zonal centrifugation for the analysis of the complexome, several considerations have to be taken.

First, a solute with which the gradient should be formed has to be chosen to separate the cellular content. It should be chemically and biologically inert, water-soluble and give solutions of high density and low viscosity. Glycerol and sucrose both fulfill these criteria and have been in use for rate zonal centrifugation of biological molecules since its inception (Brakke, 1961). They further stabilize proteins

and with that complexes, protecting them from denaturation during centrifugation (Gekko and Timasheff, 1981a,b; Lee and Timasheff, 1981; Timasheff, 1993). For Grad-seq, glycerol is used, as a 10–40 % (w/v) linear gradient was shown to be suitable to resolve complexes up to >50S (Smirnov et al., 2016). It further simplifies the downstream analyses, especially for proteins (own personal experience).

1.2.5.6   Specific considerations for Grad-seq: gradient resolution

The resolution in a rate zonal gradient is the thickness of the resulting zones and the distance between them. Narrow zones that are well separated from each other mean high resolution, which is obviously desirable for Grad-seq. To achieve this, several factors have to be considered (Brakke, 1961): Firstly, the thickness of a zone depends on the concentration of the molecules within the zone (they have to "fit"). Secondly, greater distance available for migration within the tube will allow better separation between the zones, but is naturally limited by the length of the tube. Thirdly, the used glycerol gradient is not only a gradient of density, but also one of viscosity, which is controlled by the steepness of the gradient. As can be seen from the Svedberg formula (Formula 1.1), an increase in viscosity will slow down the molecules in the front of the forming zone relative to the same molecules in the back of the zone, leading to narrower zones (Sharpe, 1988). At the same time, however, the decrease in sedimentation rate caused by the increasing viscosity leads to worse separation between zones (Brakke, 1961). Lastly, the temperature at which the centrifugation is performed influences the viscosity of the gradient and therefore the points just mentioned.

For Grad-seq, a middle ground has to be used to achieve the best possible resolution, as some of the criteria cannot be changed in order to optimize resolution: The applied amount of lysate has to be enough to allow downstream analyses, especially LC-MS/MS, but it should still be as little as possible to achieve narrow zones. The optimal amount of lysate has to be found empirically. The dimensions of the gradient tube have to be commercially available and allow downstream analyses, making high volume (*i.e.*, wide) tubes unsuitable because the fraction volume will increase too much. Therefore, a rather narrow but long tube is used. Since a lysate, *i.e.*, a complex sample, is applied to the gradient to perform Grad-seq, the density and

steepness of the gradient has to allow separation of the soluble cellular complexes. As mentioned before, a 10–40 % glycerol gradient was found to perform well for this purpose (Smirnov et al., 2016). The temperature for a Grad-seq experiment has to be 4 °C in order to stabilize complexes and reduce the biological activity of the cellular content, especially RNases that might otherwise compromise the experiment.

1.2.5.7   Specific considerations for Grad-seq: lysate preparation

To allow analysis of intact complexes by Grad-seq, cell lysis has to be performed rapidly and at low temperatures in order to conserve the *status quo*. The preferred method to achieve this is mechanical lysis using glass beads, as it can be performed at 4 °C and takes only tens of seconds to a few minutes. The duration and strength of the mechanical lysis has to be optimized empirically for each organism to prevent complex dissociation, which can be controlled for using conserved complexes such as 6S RNA-RNAP (Wassarman and Storz, 2000).

1.2.5.8   Application of Grad-seq

As opposed to the global complexomic methods discussed before, Grad-seq is able to study not only PPIs but also RNPs. It is performed natively and does not involve tags, making it suitable for the study of genetically intractable organisms. The high-throughput analyses of RNA and protein from the gradient fractions enables reconstruction of their migration patterns within the gradient (Figure 1.3). For instance, ribosomal RNAs (rRNAs) are expected to show congruent profiles with their respective ribosomal proteins, indicating intact complexes.

The main strength of Grad-seq is the global comparative analysis of these profiles with the help of dimension reduction methods such as principle component analysis or t-distributed stochastic neighbor embedding. Especially for RNAs, this is a unique feature of the method and allows the classification of RNAs into groups of transcripts with similar gradient profiles, suggesting they are involved in similarly organized complexes (Smirnov et al., 2017a). In *Salmonella*, for example, two major classes of sRNAs were found (Smirnov et al., 2016). One mostly contained sRNAs known to interact with the global RBP Hfq, which crucially was also co-migrating in the same fractions, indicating stable interaction. The other class contained sRNAs

whose functions were largely unknown. Following the "guilt-by-association" logic, this second class was investigated for common protein binders using the MS2 pull-down approach (see 1.2.2.2), leading to the identification of several candidates for each sRNA. However, only ProQ was pulled down by all sRNAs and at the same time showed highly correlating gradient profiles, which made it a strong candidate for a new global RBP. Indeed, subsequent protein-centric analyses (see 1.2.2.1) validated the MS2 results, establishing ProQ as a true global RBP (Bauriedl et al., 2020; Holmqvist et al., 2018; Melamed et al., 2020; Smirnov et al., 2016).

Given the separation of RNA along the gradient, Grad-seq is able to discriminate between functional species deriving from the same transcript. This is of particular interest in the case of sRNAs that need to be processed from 5′ or 3′ untranslated regions (UTRs) of mRNAs: The functional sRNA is expected to migrate away from the ribosomal fractions and might show interaction with an RBP like Hfq, whereas the full-length mRNA is expected to be translated, showing its peak abundance at the 70S ribosomes (Smirnov et al., 2017a).

At the same time, Grad-seq can also be used to investigate PPIs. While RNAs can be classified based on the binding of many RNAs to the same hub protein, this behavior is not expected for proteins. Still, there are several ways the MS data can be used to guide the analysis of complex involvement of proteins. For example, genes of subunits of a complex are often organized in an operon within the same mRNA (Wells et al., 2016), whose translational regulation leads to expression of the subunits proportional to their stoichiometry (Burkhardt et al., 2017; Li et al., 2014). Grad-seq can take advantage of this relationship by searching known operons for proteins of interest that show highly correlating migration profiles, indicating complex formation. The ever-cheaper cost for RNA-seq has proven particularly helpful for this kind of analysis, as more and more studies reveal the global operon structures for many different bacteria such as *E. coli* (Conway et al., 2014) or *S. pneumoniae* (Warrier et al., 2018).

Similarly, proteins of interest can also be investigated according to their general sedimentation behavior. To do so, the sedimentation coefficient of a protein of interest can be estimated according to its molecular weight (Erickson, 2009). If a protein in a Grad-seq experiment runs at considerably higher S values than estimated, it is

likely that it is involved in a complex. As for the RNP discovery discussed above, downstream analyses have to be performed in this case in order to accurately pinpoint PPIs of the protein in question.

1.2.5.9 Limitations of Grad-seq

As already hinted on, Grad-seq has several limitations, which are mostly represented by the achievable resolution of the method. The length of the tube, in which the gradient is formed in, as well as the amount of the lysate that needs to be applied, limit the resolution. Since the tube length cannot be increased indefinitely, an alternative would be to use higher volume gradients by using the SW 28 Ti or SW 32 Ti rotors, which can be operated with gradients of up to 38 ml volume compared to the SW 40 Ti rotor normally used for Grad-seq. This would allow sharper zones due to a thinner lysate layer and with that fractionation into more fractions, but has the downside of diluting the material, meaning that downstream analyses would be more difficult to perform. In line with this, lower amounts of applied lysate would generally be favorable for the resolution of the experiment but might also lead to sensitivity issues downstream, especially for the MS analysis. A different approach could be to replace the gradient by a different separation method like SEC or ion exchange chromatography, which, however, have other limitations such as non-linear separation in the beginning and at the end of the column (Hong et al., 2012).

The centrifugation time needed for a standard Grad-seq experiment is 17 h, which means that only stably associated complexes (*i.e.*, the predominant form of an RNA or protein) have a good chance of surviving the protocol. Consequently, transient or short-lived interactions cannot be captured unless stabilizing measures such as crosslinking are considered. Membrane-associated complexes are also difficult or impossible to capture using Grad-seq since they tend to be poorly soluble, which could be improved by the use of detergents (Carlson et al., 2019).

Finally, the sensitivity of the RNA-seq and MS analyses limit the detection of present molecules in a Grad-seq experiment and therefore the reconstruction of their sedimentation profiles. Especially for the RNA-seq data, this sometimes leads to artificial peaking of certain transcripts or even whole fractions, which makes manual

adjustment necessary even though a complex RNA spike-in is used for normalization (see 2.3.2).

1.3 Investigating the complexomes of two distant bacterial model organisms

In the present doctoral thesis, Grad-seq has been used to study the ensemble of cellular complexes in the bacterial model organisms *Escherichia coli* K-12 MG1655 and *Streptococcus pneumoniae* TIGR4, which represent important members of the Gram-negative and Gram-positive clades of bacteria, respectively. *E. coli* was used in a first step to optimize and improve the methodologies employed for the present research. In a second step, both organisms were used separately to obtain complexomic resource datasets, which were finally analyzed in a third step in order to explore potential new complexes.

1.3.1 *Escherichia coli*

E. coli is a Gram-negative, rod-shaped, facultative anaerobic, motile γ-proteobacterium. It generally is a commensal member of the gut microbiota of many animals including humans, but some serotypes have acquired virulence factors via horizontal gene transfer enabling them to infect various niches such as the gut, blood or bladder. The genomes of *E. coli* strains are extremely diverse, comprising a pangenome of >43,000 genes with only ∼870 (∼2 %) of them being considered part of the core genome (Yang et al., 2019). *E. coli* carries these genes on a single chromosome and, depending on the strain, none to several plasmids. The most studied strain, which is also the strain that was used in the present doctoral thesis, is *E. coli* K-12 MG1655. It is often called the "workhorse of microbiology" and is probably the organism we have the most complete understanding of. MG1655 is a derivative of a non-pathogenic K-12 strain isolated in the 1920s that was cured of the bacteriophage lambda and the F plasmid. Due to its non-pathogenicity, fast growth and easy handling, it became the most popular bacterium to study and was also one of the first to be fully sequenced (Blattner et al., 1997).

In addition to its contributions to the understanding of fundamental biological processes, *E. coli* was the model used for the identification of sRNA-mediated regulation in bacteria, the first example being MicF (Inouye and Delihas, 1988). While the first sRNAs were mostly found by chance (Wassarman et al., 1999), later studies used computational predictions in order to discover an ever-increasing amount of sRNAs (Argaman et al., 2001; Chen et al., 2002; Wassarman et al., 2001). Moreover, *E. coli* Hfq was used to describe it as a global sRNA-interacting RBP needed for the function of many sRNAs by facilitating base pairing to their mRNA targets (Babitzke et al., 2019; Gorski et al., 2017; Holmqvist and Vogel, 2018). Apart from Hfq, *E. coli* also contains other regulatory RBPs such as the translation-inhibiting CsrA, the global RNA chaperone ProQ and cold shock proteins (Babitzke et al., 2019; Holmqvist and Vogel, 2018). The interactomes of these RBPs have been characterized using several of the techniques described in 1.2 (Holmqvist et al., 2018; Melamed et al., 2020; Potts et al., 2017). Therefore, we have a good understanding of RBPs and their interacting RNAs in *E. coli*.

Next to the study of RNA-based regulation, *E. coli* has also been investigated with regards to its complexome. These studies investigated PPIs using global AP/MS (see 1.2.3) (Arifuzzaman et al., 2006; Babu et al., 2018; Butland et al., 2005; Hu et al., 2009), global Y2H (see 1.2.3) (Rajagopala et al., 2014) and 2D-MS (see 1.2.4.1) (Diéguez-Casal et al., 2014; Lasserre et al., 2006; Pan et al., 2010, 2011), leading to comprehensive datasets of the global interactome of proteins.

The ongoing interest in the biology of *E. coli* has led to the accumulation of a vast number of databases that are specific to this organism. One of the most important is EcoCyc, which aims to be a complete source of information about the molecular catalog of *E. coli*, *i.e.*, its genes, their functions, genetic context, regulation, *etc.* (Keseler et al., 2017). RegulonDB is a resource about gene regulation (Santos-Zavaleta et al., 2019), GenExpDB is a gene expression database comprised of a large number of microarray studies[1] and EcoliWiki is a community-driven *E. coli* encyclopedia (McIntosh et al., 2012). *EcoSal Plus* takes a different approach and represents an ongoing review journal about everything concerning *E. coli* and the related

---

[1]https://genexpdb.okstate.edu/
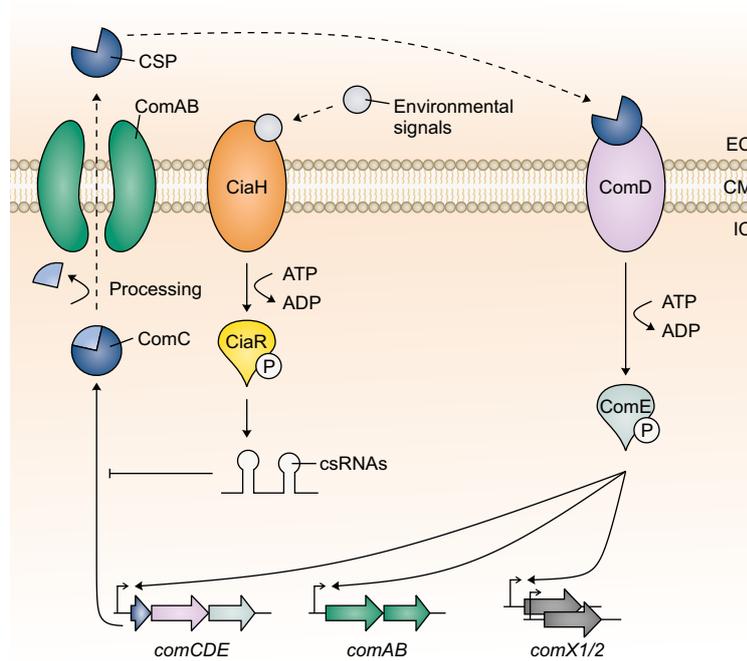
Figure 1.4: Model of the competence regulon of *S. pneumoniae*. The pneumococcal competence regulon is a quorum sensing system activated by the competence stimulating peptide (CSP). CSP induces the two-component system (TCS) ComDE, which then activates expression of the early competence loci *comAB*, *comCDE*, *comX1* and *comX2*. This results in a positive feedback loop producing ComC, which gets processed to CSP and exported by ComAB. ComX1 and ComX2 are paralogous alternative σ-factors activating the late competence genes responsible for DNA uptake and integration. CiaRH is a second TCS that is triggered by a variety of environmental signals and induces expression of the cia-dependent sRNAs (csRNAs), which are regulatory RNAs that inhibit competence by post-transcriptional repression of *comC* translation. EC, extracellular. CM, cell membrane. IC, intracellular. Adapted from Hör et al. (2020a).

Salmonella: from historical perspectives to genomics, from metabolism to systems biology.

The successes in the investigation of *E. coli* biology notwithstanding, there still is a gap in the understanding of its global complexome due to the omission of RNPs in past studies, thereby making *E. coli* an interesting model for the application of Grad-seq. Furthermore, comparison to the previously published Grad-seq dataset on the closely related *Salmonella* (Smirnov et al., 2016) would allow the investigation of the conservation of complex formation.

1.3.2 *Streptococcus pneumoniae*

S. pneumoniae, the pneumococcus, is a Gram-positive, facultative anaerobic, non-motile coccus in the phylum of Firmicutes. Until its renaming, it was called "*Diplococcus pneumoniae*" because it is usually found in pairs of two cells. The pneumococcus is one of the major bacterial pathogens, causing a wide range of infectious diseases such as otitis media, sinusitis, sepsis, meningitis and pneumonia. In children below the age of five, it annually causes more than one million deaths—more than any other infectious disease (Henriques-Normark and Tuomanen, 2013; O'Brien et al., 2009).

In contrast, it is a commensal colonizer of the upper respiratory tract in the majority of children and a small percentage of adults, which is the reservoir allowing it to spread between people. Via unknown mechanisms, events like respiratory tract infections can trigger the pneumococcus to switch toward becoming an opportunistic pathogen, disseminating to the sinuses, lungs, blood and eventually the brain (Henriques-Normark and Tuomanen, 2013). The pneumococcal capsule is its major virulence factor comprising >90 different serotypes, which makes it difficult to contain this pathogen via vaccination (Geno et al., 2015). This is further complicated by its ability to switch between serotypes.

Similar to *E. coli*, the pneumococcus has a very plastic genome with up to 13,000 genes making up its pan-genome. Of these, only 400–1,100 genes are considered part of the core genome (Hiller and Sá-Leão, 2018). Importantly, 224 pneumococcal genes are associated with virulence, separating it from commensal streptococci (Kilian and Tettelin, 2019). To encode for all these genes, the pneumococcus carries a single chromosome, which was one of the first bacterial genomes to fully be revealed by sequencing (Tettelin et al., 2001), and no plasmids. An important factor for this genomic plasticity is the pneumococcus' natural competence, allowing the uptake and integration of foreign DNA (Muschiol et al., 2019; Salvadori et al., 2019). This ability was the foundation to some of the most important studies in the field of biology, leading to the discovery of genes being made of DNA (Avery et al., 1944; Griffith, 1928) and paving the way to a new field called "molecular biology".

Pneumococcal competence is a quorum sensing system activated by secretion of competence stimulating peptide (CSP), which then activates the two-component

system (TCS) ComDE leading to expression of the early competence operons *comAB* and *comCDE*. In a second wave, late competence genes, which are required for the binding and uptake of DNA, get expressed via the paralogous alternative σ-factors ComX1 and ComX2 (Figure 1.4) (Shanker and Federle, 2017).

The competence regulon is currently also the only pathway in the pneumococcus known to involve sRNAs, namely csRNA1–5 (cia-dependent small RNAs) and srn206 (Acebo et al., 2012; Halfmann et al., 2007; Laux et al., 2015; Marx et al., 2010; Schnorpfeil et al., 2013; Wilton et al., 2015). The expression of the highly similar csRNAs is triggered by the activation of the TCS CiaRH and leads to post-transcriptional downregulation of the CSP-precursor mRNA *comC*, thereby inhibiting competence and increasing the signaling threshold needed for CSP production. Mutational studies could show that base pairing between csRNAs and their targets is needed for function, but it remains unclear whether an additional factor like an RBP is involved in the regulation. Similarly, srn206 is able to downregulate competence, which might happen through targeting of *comD* (Acebo et al., 2012).

Apart from the csRNAs and srn206, several studies identified a total of ∼70 experimentally verified sRNAs and an additional ∼100 non-verified sRNAs (Acebo et al., 2012; Mann et al., 2012; Sinha et al., 2019; Slager et al., 2018; Tsui et al., 2010; Warrier et al., 2018). The mechanisms and targets of these sRNAs remain unknown, while a transposon sequencing study showed the involvement of many sRNAs in the regulation of virulence, to a degree that single knock-outs of some of them led to a complete loss of pathogenicity in mouse infections (Mann et al., 2012). Interestingly, all five csRNAs were among the sRNAs whose disruption via transposons caused reduced fitness in the colonization of mice, highlighting the importance of these sRNAs on the one hand and of the competence regulon on the other hand (Mann et al., 2012).

The pneumococcus does not possess any homologs of CsrA, Hfq or ProQ (Tettelin et al., 2001). Together with the large number of identified sRNAs, this raises the question whether there is an overlooked regulatory RBP that is involved in the molecular mechanism of post-transcriptional regulation in the pneumococcus, thereby making it an excellent Gram-positive model for the application of Grad-seq.

1.4   Aims of this thesis

The importance of complexes for all living organisms is well established. Many methods have been developed to analyze the interactions between protein and RNA in order to understand the targets and functions of these molecules. However, generic methods to globally and unbiasedly analyze the complexome are still lacking. This thesis aimed to improve Grad-seq as a global complexomic method in order to make it applicable to any bacterial organism and provide resources for two important model bacteria. Therefore, three consecutive tasks were studied:

1. Optimization of the existing Grad-seq protocol to make it more generic and robust.

2. Application of Grad-seq to two distant bacterial species—*E. coli* and *S. pneumoniae*—to provide global complexome resources for both.

3. Analysis of the datasets with a focus on the identification of putative new complexes and interactions.

The successful establishment of these datasets would greatly benefit the bacterial community in studying their own molecules of interest.

Chapter 2

Grad-seq of *Escherichia coli*

Parts of the results obtained in this chapter were performed in collaboration with the following people:

- Silvia Di Giorgio (ZB MED, Cologne), Prof. Dr. Konrad U. Förstner (ZB MED, Cologne) and Dr. Jens Vanselow (Rudolf Virchow Center, Würzburg) performed parts of the high-throughput data analyses.

2.1 Optimization of the Grad-seq protocol

In its original application, Grad-seq was optimized for the γ-proteobacterium *Salmonella* (Smirnov et al., 2016). Conceptually, Grad-seq should, however, be applicable to all bacteria, provided a soluble lysate of intact cellular complexes can be obtained. In order to maximize the applicability of Grad-seq, almost all steps of the protocol will be addressed and optimized in the following sections, using *E. coli* as a model organism.

2.1.1 Optimization of the loading volume

In the original Grad-seq protocol applied to *Salmonella* (Smirnov et al., 2016), 200 $OD_{600\,nm}$ of cells ($\sim 200 \times 10^9$ cells) was lysed and subjected to gradient centrifugation. This large amount was necessary in order to allow the downstream analyses of the protocol. Since then, RNA-seq and MS technologies have become much more
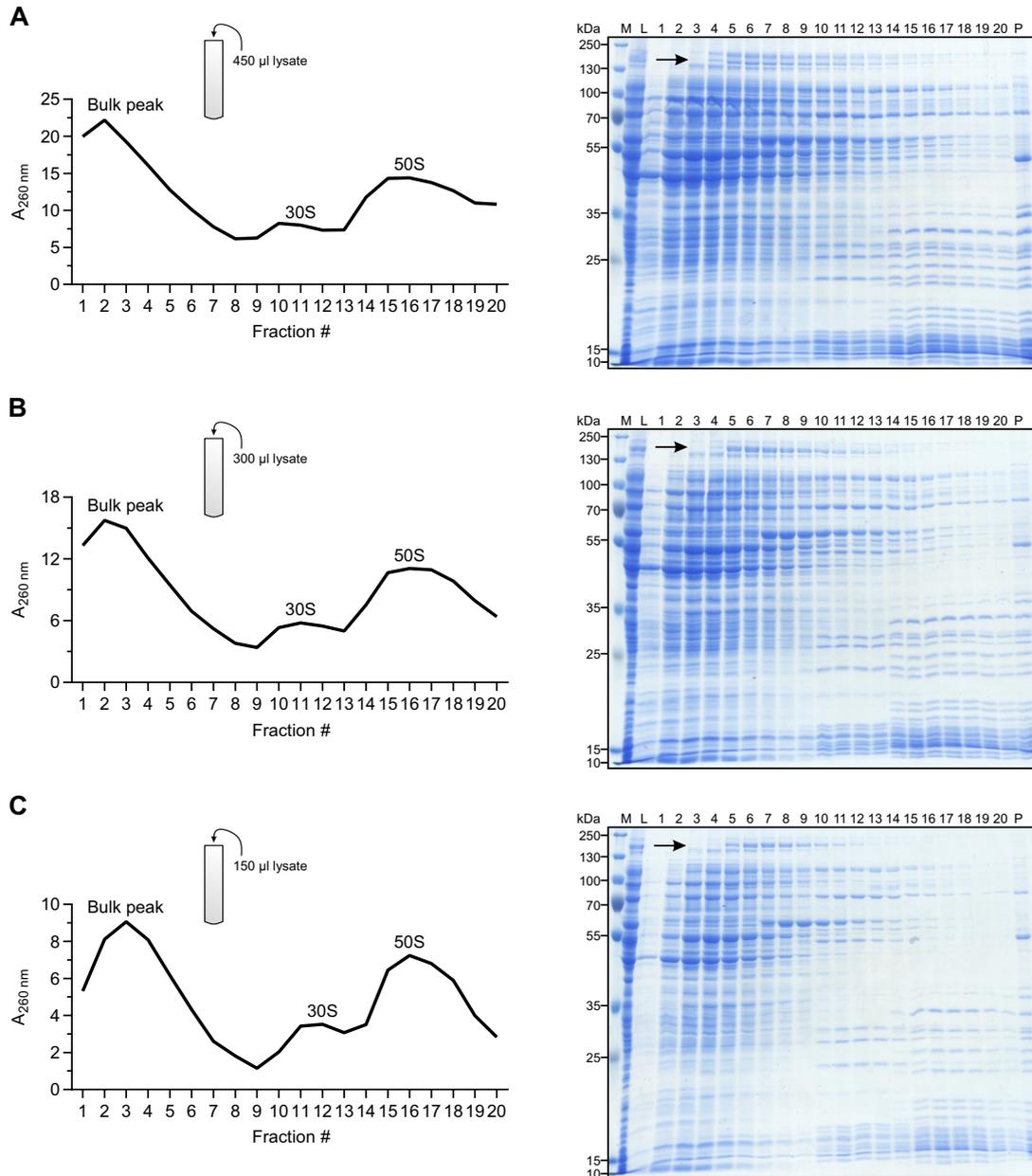
Figure 2.1: Optimization of the loading volume. 450 µl (**A**), 300 µl (**B**) and 150 µl (**C**) of identically prepared *Salmonella* lysates were loaded on 10–40 % glycerol gradients and separated via ultracentrifugation. The general gradient profiles were monitored using UV measurements (left panels) and Coomassie-stained gels (right panels). Lower loading volumes result in better separation of the cellular content. RNA polymerase subunits RpoB and RpoC are marked by a arrows. M, size marker. L, lysate (input control). P, pellet.

powerful. This is especially true for the sensitivity of MS detectors, allowing the detection of less abundant peptides than before. Consequently, it should be possible to lower the number of cells used for centrifugation, which should improve the resolution of the experiment (see 1.2.5.6). As proof of principle, three *Salmonella* gradients were prepared using the original protocol, the only difference being the amounts of lysate that were layered on top of the gradients: 450 µl (Figure 2.1 A), 300 µl (Figure 2.1 B) and 150 µl (Figure 2.1 C). Importantly, the same number of cells was used for lysis in the same volume, *i.e.*, the disruption of the cells was not different between the experiments. The resolution of the gradients was monitored using $A_{260\,nm}$ measurements of the gradient fractions (Figure 2.1, left panels) and Coomassie-stained SDS-PAGE of their protein content (Figure 2.1, right panels).

As expected, the $A_{260\,nm}$ values representing the amount of cellular material according to the absorbance of nucleic acids decreased with lower lysate volumes. It was further evident that the three major peaks representing the bulk (small RNPs, DNA, free RNA) and the 30S and 50S ribosomal subunits became narrower and more defined when less volume was loaded. The Coomassie-stained SDS-PAGE analyses confirmed these observations: The more lysate was used, the more protein was visible on the gel. In agreement with the UV profiles, the sedimentation profiles of the proteins became better resolved with less input material, as can easily be appreciated by following the migration of RNAP within the gradients (arrows). While the β- and β′-subunits were detected from fractions 3–20 when 450 µl was loaded, they were only visible in fractions 4–14 with 150 µl loading volume. Importantly, not only was the observed RNAP zone narrower, the peak abundance (fraction 5–7) of the proteins was also better resolved, which should improve the identification of complex formation by analysis of congruent migration patterns. These results show that the resolution of Grad-seq can be improved by the application of lower volumes of lysate.

2.1.2 Optimization of lysis conditions

The optimal lysis conditions for a bacterium of interest have to be found empirically. They should be performed in a way that allows fast disruption of the bacteria without damaging the existing complexes (see 1.2.5.7). Originally, a Retsch MM 400
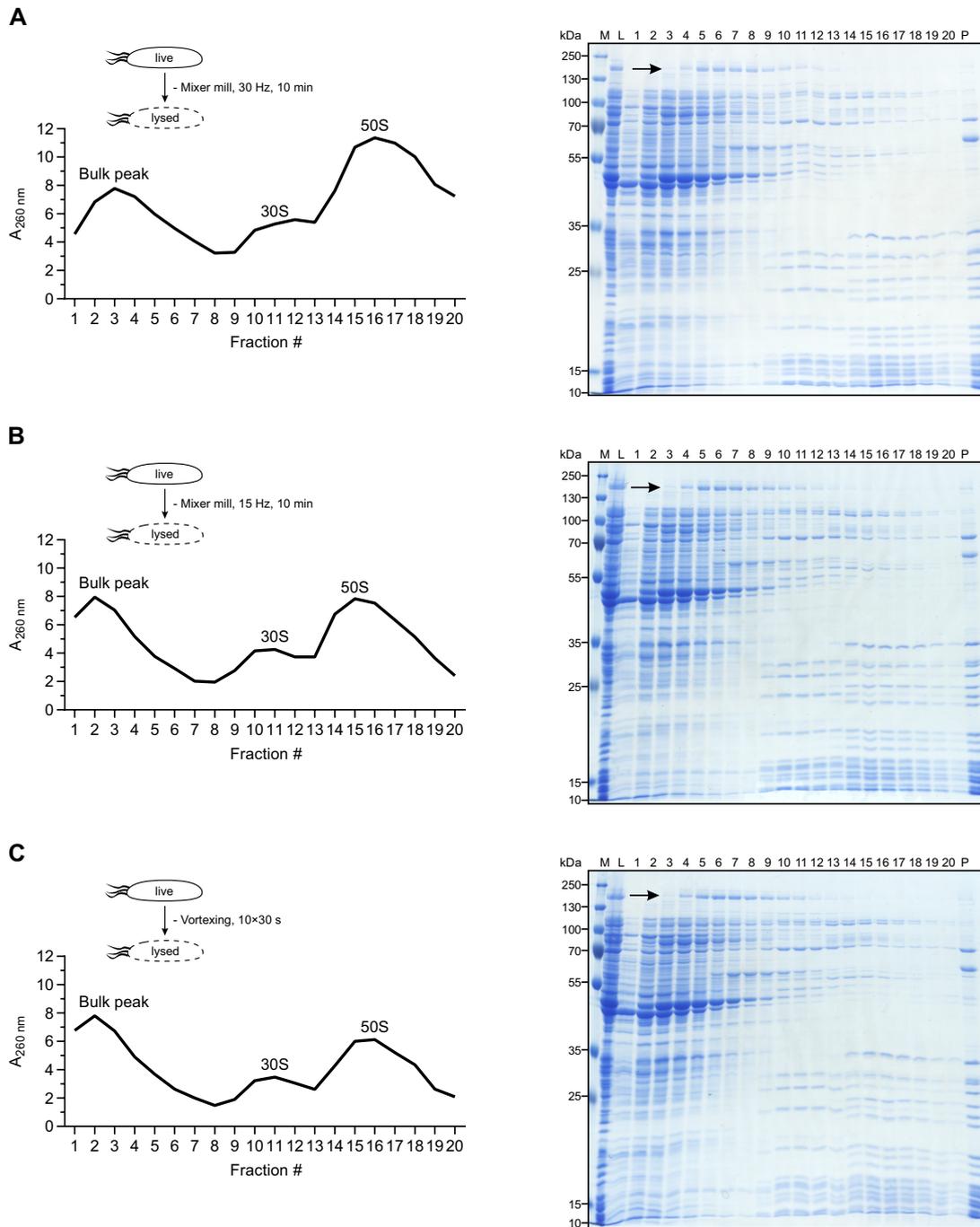
Figure 2.2: Optimization of lysis conditions. *E. coli* lysates were prepared by disruption using a mixer mill for 10 min at 30 Hz (**A**) or 15 Hz (**B**) or by 10 cycles of vortexing for 30 s followed by cooling on ice for 15 s (**C**). The lysates were separated on 10–40 % glycerol gradients via ultracentrifugation. The general gradient profiles were monitored using UV measurements (left panels) and Coomassie-stained gels (right panels). Weaker disruption leads to sharper profiles. RNA polymerase subunits RpoB and RpoC are marked by a arrows. M, size marker. L, lysate (input control). P, pellet.

mixer mill was used to disrupt *Salmonella* cells for 10 min at a frequency of 30 Hz. To investigate whether these conditions are also suitable for the closely related *E. coli*, three different lysis conditions were tested: The original protocol (Figure 2.2 A), a modified version of the original protocol reducing the frequency to 15 Hz (Figure 2.2 B) and a vortexing method used in *E. coli* gradient experiments that first showed the interaction between 6S RNA and RNAP (Figure 2.2 C) (Wassarman and Storz, 2000). The results were monitored by measuring the UV profiles of the gradients and the position of RNAP on SDS-PAGE (arrows), which should change its position if dissociated.

The UV profiles revealed that, as expected, the modified Retsch method and the vortexing method caused weaker disruption, as indicated by lower $A_{260\,nm}$ values. Interestingly, the bulk peak strongly increased with these milder lysis methods and more closely resembled profiles obtained from *Salmonella* gradients (Figure 2.1). The SDS-PAGE analyses did not show any obvious differences between the lysis methods and RNAP peaked in fractions 5–7 in all cases. Since milder lysis conditions should generally favor the stability of complexes, the vortexing method was ultimately chosen as the best lysis method for *E. coli* because it allows cooling throughout the process of lysis, whereas the mixer mill warms up during the 10 min of lysis.

2.1.3   Optimization of RNA extraction

When applied to *E. coli*, the original protocol's hot phenol RNA extraction caused several unexpected issues. First, after RNA precipitation, the pellets were often insoluble leading to loss of RNA. Second, RNA from fractions around the 30S subunit was sometimes difficult to extract at all, which also led to loss of RNA. To investigate alternative RNA extraction methods, three different methods were compared: the original hot phenol method (Figure 2.3 A), hot phenol extraction followed by chloroform extraction (Figure 2.3 B) and phenol/chloroform/isoamyl alcohol (P/C/I) extraction (Figure 2.3 C). As before, the original hot phenol method showed loss of RNA around the 30S subunit (fractions 9–12), which inexplicably was not always apparent in a EtBr-stained RNA gels. If followed by chloroform extraction, the method performed even worse and the majority of RNAs from the ribosomal fractions was lost. Gel analysis showed that 5S rRNA was still extracted with similar efficiency
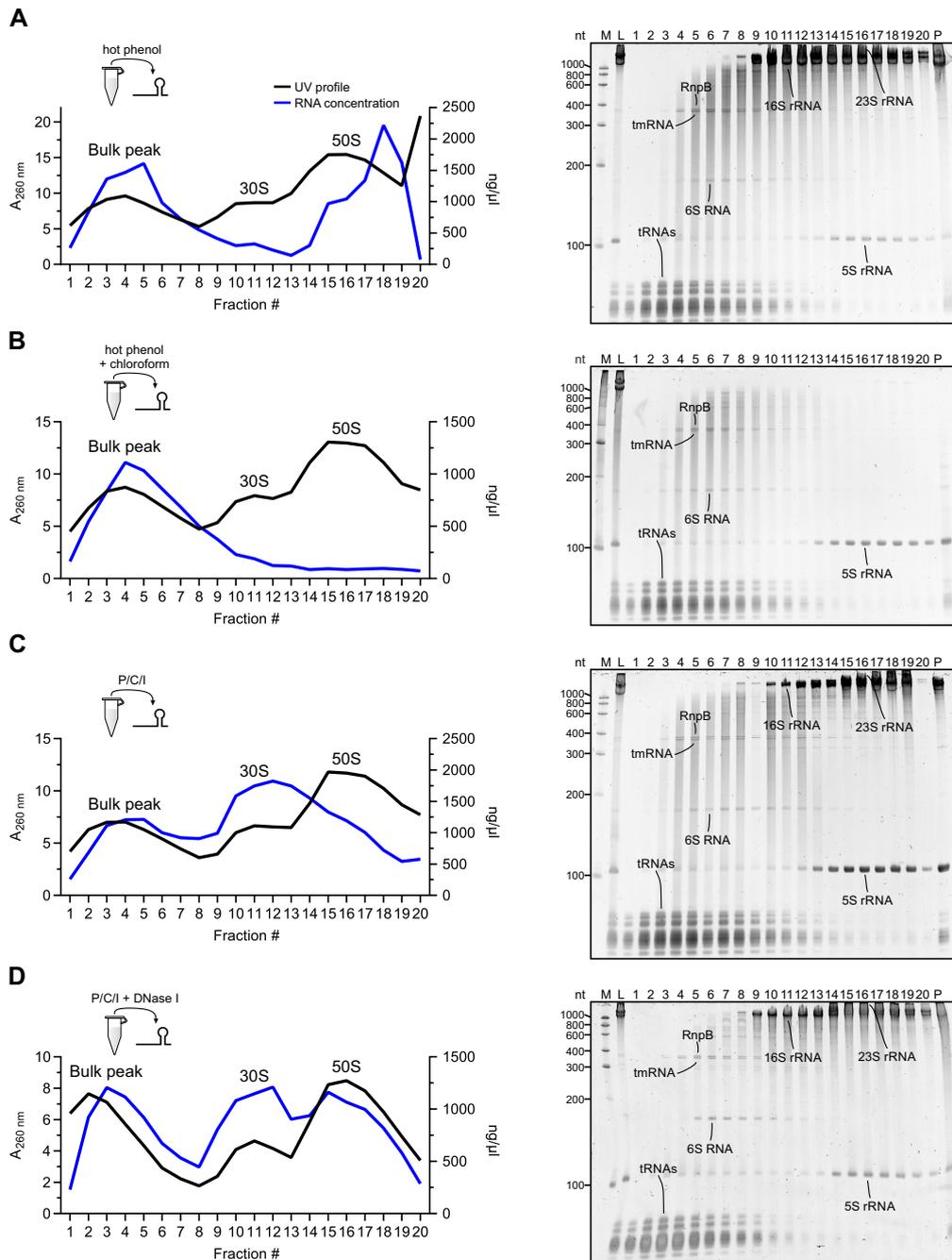
Figure 2.3: Optimization of RNA extraction. *E. coli* lysates prepared using the 30 Hz mixer mill protocol (Figure 2.2 A) were separated on 10–40 % glycerol gradients and the RNA content of the fractions was extracted using hot phenol extraction (**A**), hot phenol extraction followed by chloroform extraction (**B**) or phenol/chloroform/isoamyl alcohol (P/C/I) extraction (**C**). The general gradient profile was monitored using UV measurements (left panels) and the quality of the RNA was monitored using concentration measurements (left panels) as well as ethidium bromide-stained RNA gels (right panels). P/C/I extraction is the best method to extract RNA as judged by the RNA gels and was therefore further optimized by adding vortexing steps and digestion by DNase I (**D**). Lysis in (**D**) was performed using the vortexing method (Figure 2.2 C). M, size marker. L, lysate (input control). P, pellet.

(compare intensities of 5S rRNA and tRNA bands in Figure 2.3 A and B) and that almost all of the 16S and 23S rRNAs was lost. While removal of rRNA from the gradient should be of no concern, it cannot be excluded that other RNA species got lost as well, making this method a poor choice. In contrast to the original hot phenol method, P/C/I extraction led to improved RNA recovery around the 30S subunit but decreased recovery around the 50S subunit. As for the second method, this was mainly caused by a loss of 23S rRNA (compare intensities of 5S rRNA and 23S rRNA bands in Figure 2.3 A and C). Crucially, P/C/I extraction dramatically improved the solubility of the resulting RNA pellets.

Since P/C/I extraction was the most promising, vortexing steps were introduced in order to improve the recovery of RNA around the 50S subunit (Figure 2.3 D). Moreover, DNase I digestion was added to the protocol since its omission caused issues during trial runs of library preparation for the RNA-seq part of Grad-seq (data not shown). This revealed the sheer amount of DNA that was recovered without DNase treatment, as shown by the absence of the usual smears in fractions $\sim$3–13. This method of RNA extraction turned out to be extremely robust; it was successful in the RNA extractions of Grad-seq experiments of many different bacterial species (*e.g., S. pneumoniae*; see 3). Together, these results show that P/C/I extraction followed by DNase I digestion is the preferred way to isolate RNA from glycerol gradients in a Grad-seq experiment.

2.2   Overview of gradient sedimentation

After establishing the new protocol for *E. coli* (see 2.1), final gradients were prepared for Grad-seq using *E. coli* cultures grown to early stationary phase (OD$_{600\,nm}$ of 2.0). The coarse-grained analysis by RNA gels and SDS-PAGE indicated high reproducibility between the replicates on RNA and protein level (Figure 2.4 A and B). Further, some of the major RNPs such as 6S RNA-RNAP and the ribosomal subunits clearly showed congruent profiles of their subunits, suggesting these complexes remained intact during sample preparation. Since the gradient conditions had been chosen in a way that everything larger than a 50S subunit was pelleted, proteins and rRNA from both ribosomal subunits, representing whole 70S ribosomes, were found
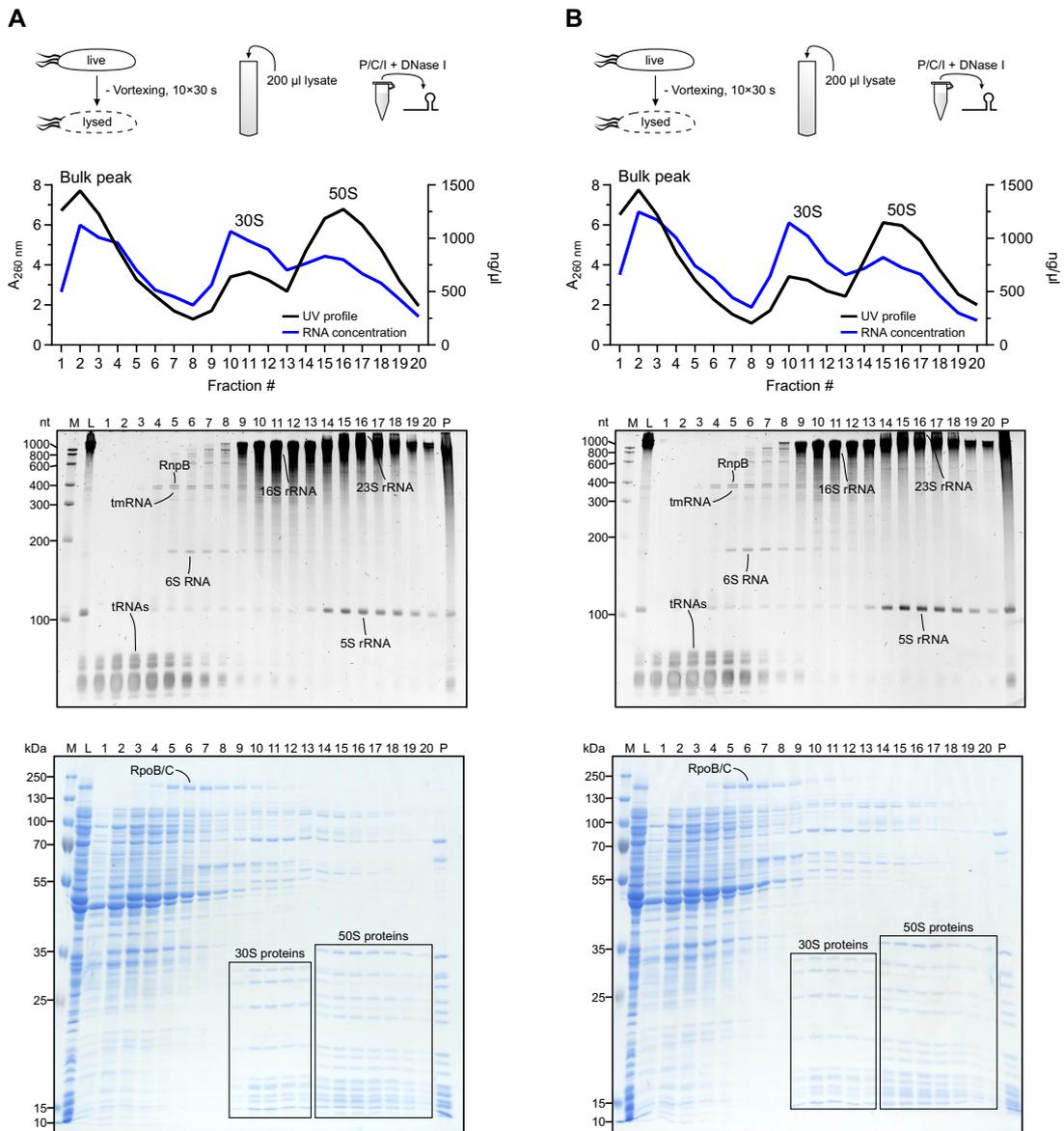
Figure 2.4: Overview of gradient sedimentation. (**A**) Using the optimized Grad-seq protocol, an *E. coli* gradient was run and its general profile monitored by UV measurements (top). The RNA sedimentation and quality was monitored using concentration measurements (top) and ethidium bromide-stained RNA gel analysis (middle). The protein sedimentation and quality was monitored using SDS-PAGE followed by Coomassie staining (bottom). The RNA and protein components of the ribosomal subunits show congruent sedimentation, indicating intact complexes. (**B**) Replicate of (**A**) showing identical results. M, size marker. L, lysate (input control). P, pellet.

in the pellet fraction. While tRNAs were mostly detected in the upper fractions of the gradient, tmRNA and RNase P RNA showed very similar sedimentation profiles, which is reminiscent of their discovery in a 10S fraction of *E. coli* resulting in their original nomenclatures as 10Sa and 10Sb RNA, respectively (Karzai et al., 2000).

To get a more detailed insight into the distribution of especially ncRNAs, northern blots were performed. Again, replicates showed consistent profiles for the tested RNAs and confirmed the positions of stable RNAs such as tRNAs, rRNAs, tmRNA and RNase P RNA as observed in the RNA gel (Figure 2.5 A and B). Together with western blotting for the β-subunit of RNAP (RpoB) and the major σ-factor σ^{70} (RpoD), the positioning of 6S RNA provided further strong evidence for an intact RNAP complex (Figure 2.5 C) (Wassarman and Storz, 2000). As previously shown for *Salmonella* (Smirnov et al., 2016), 4.5S RNA, the RNA subunit of the signal recognition particle (Akopian et al., 2013), was present in low molecular weight fractions, indicating its association with nascent peptides is not stable under the present conditions.

The Hfq-dependent sRNAs GcvB, MicA, CyaR, RprA and Spot 42 all showed peaks around fraction 5 with most of them having a surprisingly high abundance in the pellet fraction, which was not the case for ChiX (Figure 2.5 A and B). In contrast, mRNAs (*cspE*, *rpoA* and *rplU*) exhibited strong peaks in the pellet, indicating active translation and 70S ribosome association. *cspE* further showed a strong signal throughout the gradient, suggesting lower levels of translation in the present conditions, which, however, is surprising given the high protein copy numbers of CspE independent of the growth conditions (Li et al., 2014; Schmidt et al., 2016).

To validate the surprising sedimentation behavior of some of the ncRNAs, polysome gradient analysis was performed (Figure 2.5 D). As observed in the glycerol gradients, 6S RNA, ChiX and CsrB did not show any association with the polysomes. Strikingly though, both GcvB and Spot 42 were detected in polysomal fractions, thereby confirming the previous observations. Spot 42 has a conserved ORF and Shine-Dalgarno (SD) sequence (Yano, Chao and Vogel, unpublished; Gisela Storz, personal communication), which might explain its polysome association. In comparison to these Hfq-dependent sRNAs, the sedimentation of *cspE* could not be confirmed. It only showed a slight abundance in the bulk of the gradient, with the
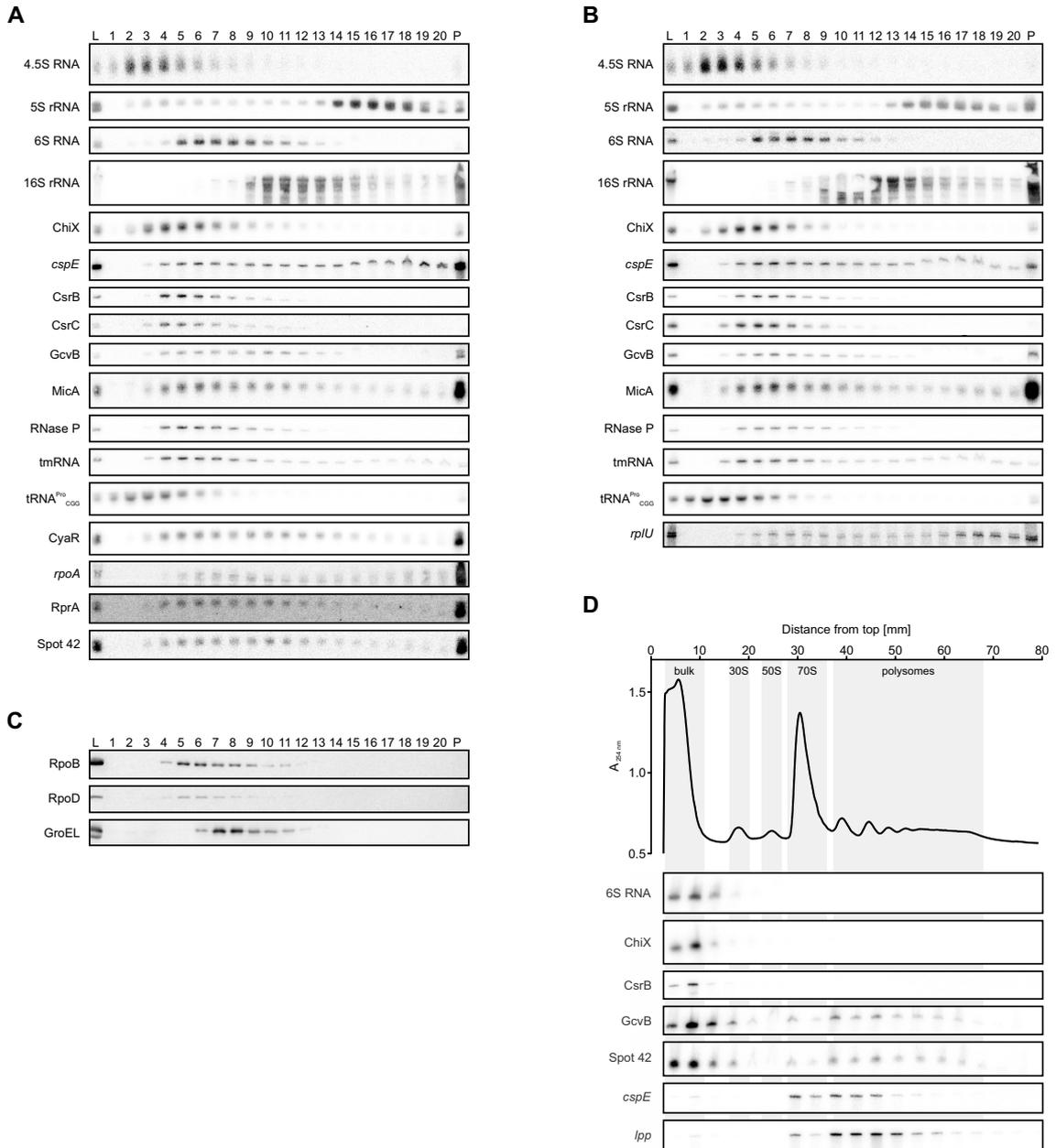
Figure 2.5: Sedimentation of ncRNAs. (A, B) Northern blots of two replicates (compare to Figure 2.4) reveal different sedimentation profiles for different ncRNAs. Interestingly, many sRNAs show their peak abundance in the pellet. (**C**) Western blots reveal the sedimentation of the RNAP subunits RpoB and RpoD as well as the chaperone GroEL. (**D**) Sucrose polysome gradient analysis verifies the polysome-association of certain sRNAs observed in (**A**) and (**B**). L, lysate (input control). P, pellet.

majority of its copies being associated to the polysomal fractions, as is typical for a translated mRNA like the control *lpp*. Since the Grad-seq protocol uses a lower concentration of Mg^{2+} ions, 70S ribosomes perhaps were destabilized (Ron et al., 1968) and therefore release of the *cspE* mRNA occurred. Together, these results demonstrate that the optimized Grad-seq protocol is reproducible and that its resulting sedimentation profiles represent stable *in vivo* complexes.

2.3 Optimization of the RNA-seq protocol

To further improve the Grad-seq protocol, several changes had to be introduced to the RNA-seq protocol.

2.3.1 Fragmentation

In the original application, the gradient RNA was sequenced on an Illumina HiSeq 2000 instrument (Smirnov et al., 2016). Since this platform has mostly been replaced, a switch to the NextSeq 500 instrument was performed, partially because it is available at the Core Unit SysMed at the University of Würzburg. In comparison to the HiSeq 2000, which tolerates long cDNA fragments, the NextSeq 500 requires shorter cDNA molecules of <600 bp to ensure proper cluster formation (Fritz Thümmler, Vertis Biotechnologie AG, personal communication). To obtain this length, fragmentation of the extracted gradient RNA was introduced. In addition, a gel-based size selection step was added after cDNA preparation and pooling of the samples. To compensate for the expected flood of rRNA after fragmentation, the pooling factors for each fraction were chosen according to the rRNA concentration as estimated from the UV profile.

2.3.2 Use of a complex RNA spike-in allows better normalization

A critical step in the analysis of a global Grad-seq dataset is the normalization of the read counts to an external standard. With this, differences in library size for the single fractions can be eliminated and the sedimentation profiles accurately calculated. Originally, a single, 20 nt long spike-in RNA was used for this purpose,
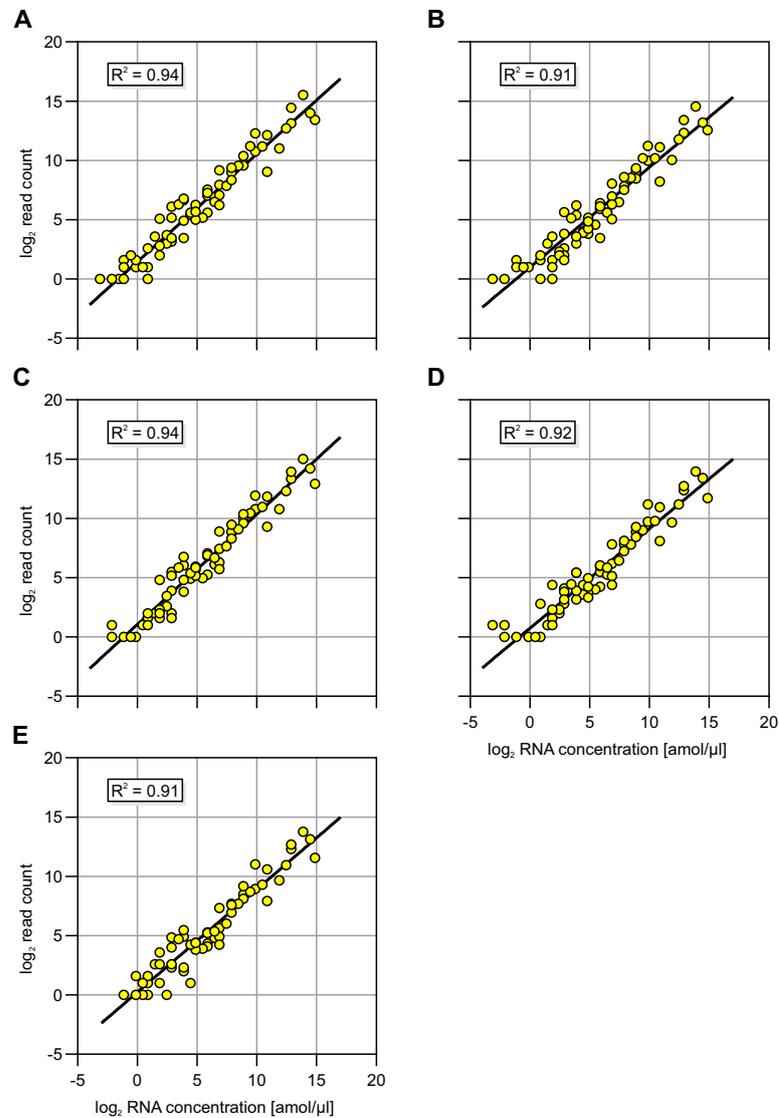
Figure 2.6: Performance test of the ERCC spike-in. ERCC spike-in was added to the purified RNA of fractions 2 (**A**), 4 (**B**), 7 (**C**), 10 (**D**) and 16 (**E**) of an *E. coli* gradient (compare to Figure 2.4). Following analysis by RNA-seq, the obtained number of reads for the spike-in transcripts were plotted against the known concentration of the corresponding spike-in transcripts. All tested fractions show excellent linear correlation between the read counts and the known concentration, indicating that the ERCC spike-in is a suitable tool for the normalization of a Grad-seq experiment. R^2 is the coefficient of determination.

which, however, was not performing fully satisfyingly, requiring additional manual normalization (Smirnov et al., 2016). The spike-in developed by the External RNA Control Consortium (ERCC) (Baker et al., 2005; The External RNA Controls Consortium, 2005) was chosen as a commercially available replacement. It is made up of 92 transcripts from 250–2,000 nt in length that span a 10^6-fold concentration range and which were counter-selected against several transcriptome databases including bacteria. Except for *B. subtilis*, no homologies to tested bacterial transcriptomes are present in the ERCC spike-in, which should be double-checked when applying Grad-seq to an understudied organism.

To test the performance of the ERCC spike-in, 5 gradient fractions (fractions 2, 4, 7, 10 and 16; compare to Figure 2.4) including the highest and lowest RNA concentrations of the gradient were chosen, the spike-in added and the samples sequenced. The obtained read counts for the spike-in transcripts should be linearly increasing with their known physical concentrations. Indeed, a very strong linear correlation ($R^2 \geq 0.91$) could be observed in the tested samples, detecting between 64–74 of the spike-in transcripts and covering a dynamic range of $\sim 10^5$ (Figure 2.6 A–E). The complexity of the ERCC spike-in, together with its large concentration span should be well-suited for the normalization of Grad-seq datasets (Risso et al., 2014), possibly eliminating the need of manual adjustments. Computational normalization is achieved by DESeq2 (Love et al., 2014), which considers the spike-in as a set of "control genes" that do not change in concentration between samples and which are used to calculate size factors for each fraction. This normalization is built-in in GRADitude[1], a tool specifically designed for the analysis of Grad-seq data (Di Giorgio, Hör, Vogel and Förstner, unpublished).

2.4 Overview of the *E. coli* Grad-seq RNA data

2.4.1 Sedimentation profiles of total RNA

Before generation of the libraries for the *E. coli* Grad-seq data set, capillary electrophoresis was performed to ensure the quality of the RNA (Appendix Figure 7.1 A), showing results comparable to the RNA gel (Figure 2.4). Surprisingly, frac-

---

[1]https://foerstner-lab.github.io/GRADitude/

A

B
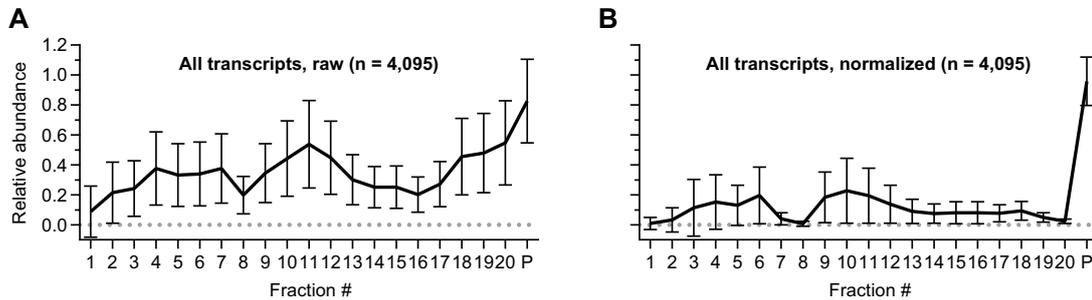
Figure 2.7: Transcript sedimentation before and after normalization. (**A**) Average sedimentation profile of all detected transcripts without ERCC-based normalization. An early peak and a peak around the 30S subunit (fraction 11) are visible. Toward the end of the gradient, the relative abundance increases. (**B**) Average sedimentation profile of all detected transcripts after ERCC-based normalization. In comparison to the profile observed in (**A**), there is no increase in relative transcript abundance at the end of the gradient. Yet, the peak in the pellet remains and overall is emphasized. Profiles are normalized to a maximum of 1. Error bars show SD from the mean.

tion 8 seemed to contain almost no RNA, which had not been observed in previous analyses. Since less RNA was applied to the electrophoresis chip, this could be caused by detection limits. After fragmentation and cDNA preparation, some strong bands appeared, which mostly belonged to abundant transcripts (tRNAs, 6S RNA, 5S rRNA) present in the corresponding fractions and which were seemingly left intact (Appendix Figure 7.1 B).

RNA-seq resulted in ∼433 million reads in total. The relative abundances of the major RNA classes were expectedly very different between the fractions: while reads deriving from tRNAs dominated the first fractions, up to ∼97 % of the reads in fractions 15 and 16 derived from rRNAs (Appendix Figure 7.1 C). However, since the pooling factors were chosen according to the estimated rRNA abundance, >1 million reads deriving from mRNAs and ncRNAs were sequenced for all fractions except fraction 2, which had only slightly less. After filtering out low abundance transcripts (row sum of <100 reads), a total of 4,095 transcripts was detected within the gradient. Before normalization using the ERCC spike-in, the average sedimentation profile of these transcripts was already reminiscent of the data obtained for *Salmonella* (Smirnov et al., 2016), showing the expected peaks in the beginning of the gradient, around the 30S subunit and in the pellet (Figure 2.7 A).

Yet, the abundance in fractions 17–20 was high, as was the overall abundance within the gradient. Normalization of the data reversed this observation, leading
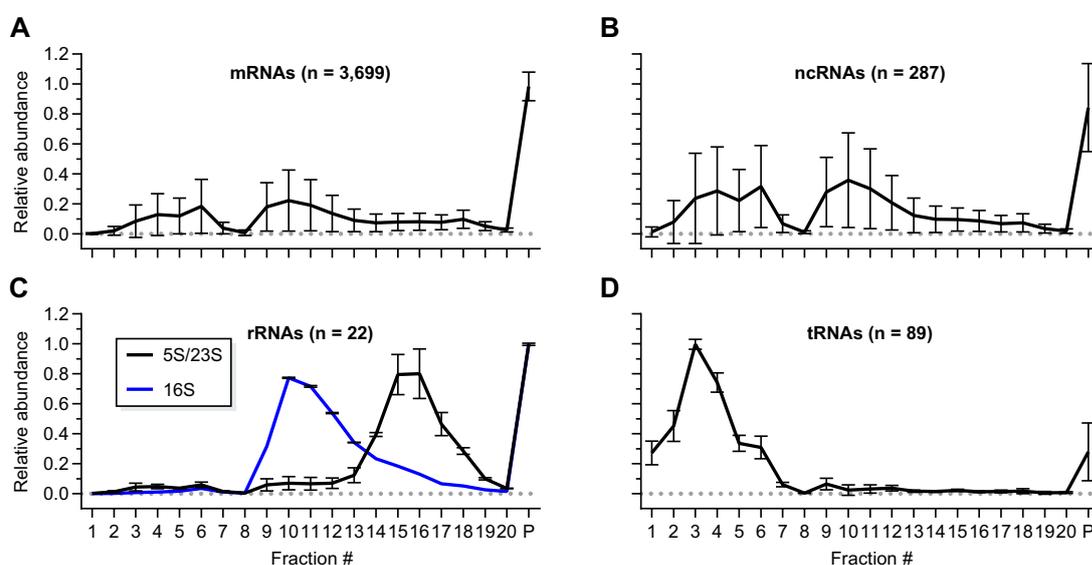
Figure 2.8: Sedimentation profiles of different RNA classes. The average sedimentation profiles of mRNAs (**A**), ncRNAs (**B**), rRNAs (**C**) and tRNAs (**D**) reveal that different RNA classes show different gradient sedimentation behavior. Profiles are normalized to a maximum of 1. Error bars show SD from the mean.

to a strong increase in the average abundance in the pellet fraction (Figure 2.7 B). Since most of the sequenced transcripts derived from mRNAs (∼90 %), a strong average association with the 70S ribosome was expected, agreeing with the profiles obtained after normalization. A small distortion of the normalized average profile was observed in fraction 6, which showed higher relative abundance than expected. At this point, no manual adjustment was applied, though, because it was not possible to judge whether the read counts in fraction 6 were too high or whether the read counts in fraction 5 were too low.

2.4.2 Sedimentation profiles of different RNA classes

To gain a more detailed overview of the RNA-seq data, the average sedimentation profiles of different classes of RNA were studied. mRNAs migrated mostly with the 70S ribosome and less with the 30S subunit and the bulk peak (Figure 2.8 A), which likely represented active translation. In contrast, ncRNAs, which often bind to one of the major RBPs (Babitzke et al., 2019; Holmqvist and Vogel, 2018; Hör et al., 2020b), showed a similar profile but with far higher abundance in the soluble fractions of the gradient (Figure 2.8 B), especially in the bulk and 30S regions, as
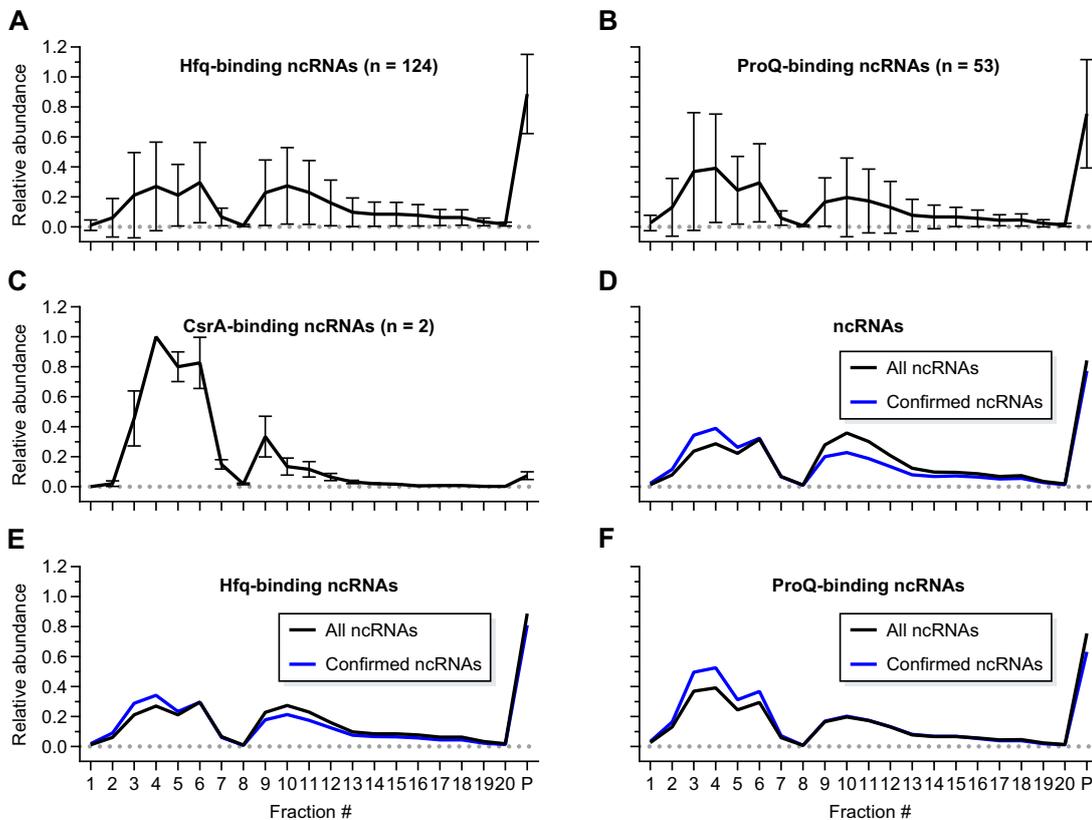
Figure 2.9: Sedimentation of different classes of ncRNAs. The average sedimentation profiles of Hfq-binding (**A**), ProQ-binding (**B**) and CsrA-binding (**C**) ncRNAs reveal that Hfq-binding ncRNAs tend to peak around fraction 4, around the 30S subunit and in the pellet fraction. The same is true for ProQ-binding ncRNAs, though their early peak is more pronounced. CsrA-binding ncRNAs on the other hand only show a strong peak around fraction 5. (**D, E, F**) When only confirmed ncRNAs are considered, the average gradient profiles tend to show higher abundance in early fractions and lower abundance around the 30S subunit as well as in the pellet, indicating falsely annotated ncRNAs are present in the list of all ncRNAs. Profiles are normalized to a maximum of 1. Error bars show SD from the mean.

observed before using northern blotting (Figure 2.5 A and B). In agreement with the gradient UV profile, the rRNAs showed peaks at the 30S and 50S regions for the 16S and the 5S/23S rRNAs, respectively (Figure 2.8 C), whereas tRNAs were almost exclusively found around fraction 3 (Figure 2.8 D).

To refine the profiles of sRNAs, they were further clustered according to their cognate RBPs (Hfq/ProQ) as identified by CLIP-seq (Michaux, Hör and Vogel, unpublished; Holmqvist et al. (2018)). As shown previously in *Salmonella* (Smirnov et al., 2016), Hfq-binding sRNAs showed peaks around fractions 4–6 (Figure 2.9 A), while ProQ-binding sRNAs peaked earlier and showed higher abundance within

the gradient (Figure 2.9 B). The CsrA-titrating sRNAs CsrB and CsrC (Romeo and Babitzke, 2018) on the other hand showed almost no presence in the pellet fraction (Figure 2.9 C), as revealed by northern blotting before (Figure 2.5 A and B). The *E. coli* MG1655 annotation version used for the analysis of the RNA-seq data contained 99 confirmed sRNAs and 188 predicted ones that have not been validated experimentally.

Looking at the differences in sedimentation behavior between all ncRNAs and only the confirmed ones (Figure 2.9 D), it became evident that the confirmed sRNAs showed higher average abundance around the bulk peak and less average abundance around the 30S subunit, as was shown for *Salmonella* sRNAs (Smirnov et al., 2016). The same was true when only Hfq- or ProQ-binding sRNAs were considered (Figure 2.9 E and F), indicating there might be false positives, such as UTRs of mRNAs, within the list of predicted sRNAs. These results demonstrate that the RNA-seq data of *E. coli* Grad-seq overall is very similar to the previous work in *Salmonella* and that the sedimentation profiles of sRNAs differs depending on the RBP they are binding.

2.4.3 Manual adjustment of the RNA-seq normalization

While the RNA-seq data fit to the previously published data, they did not fully agree with the results obtained in the laboratory. Especially the "RNA void" encountered in fractions 7 and 8 (Figure 2.7 B) was not visible on the northern blots from the same and other gradients (Figure 2.5 A and B), where the RNAs showed no sudden drop in abundance in these fractions. Inexplicably, this phenomenon could so far only be observed for *Salmonella* and *E. coli* Grad-seq data, whereas other species did not show this behavior in the RNA-seq (see, for example, Figure 3.3). In order to more accurately reflect the underlying laboratory data, the RNA-seq data was adjusted manually by employing empirically determined size factors using the northern blot data as comparison. Specifically, the spike-in normalized read counts of fraction 5 were multiplied by 1.5, whereas those of fractions 7 and 8 were multiplied by 4.5 and 28, respectively.

Afterwards, relative normalization to the maximum value for each transcript was performed as before. While the increase in Spearman's correlation was only mod-
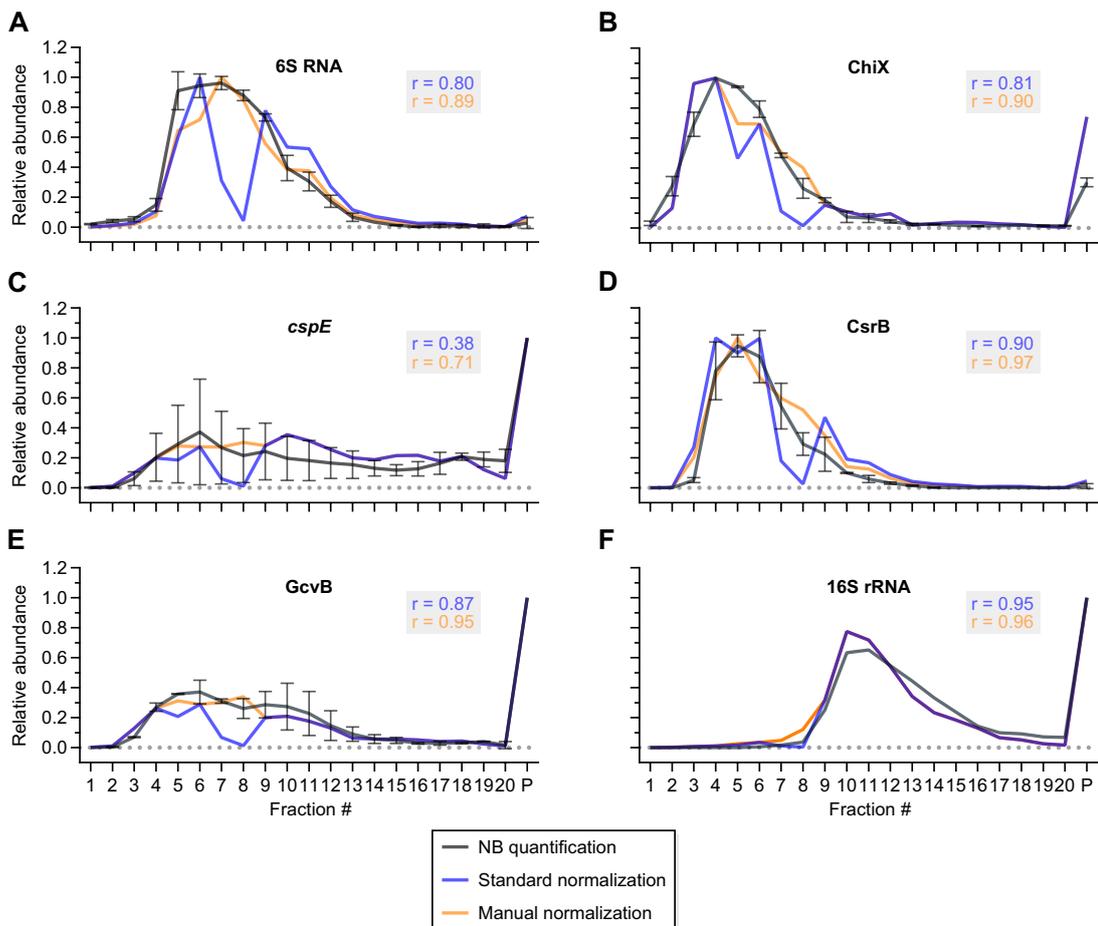
Figure 2.10: Manual adjustment of the RNA-seq normalization improves comparability to data from the laboratory. Comparison of the sedimentation profiles of 6S RNA (**A**), ChiX (**B**), *cspE* (**C**), CsrB (**D**), GcvB (**E**) and 16S rRNA (**F**) obtained from northern blots (NB, gray), RNA-seq without adjustment (blue) and RNA-seq with manual adjustment (orange). Manual normalization increases the comparability to the northern blot profiles. When available, replicate data is shown for the northern blots (n = 2). Profiles are normalized to a maximum of 1. Error bars show SD from the mean. r is the Spearman's correlation coefficient between the NB data and the standard (blue) or manual (orange) normalization.

erate, the manually adjusted RNA-seq data clearly represented the northern blot data better than the data only normalized by the ERCC spike-in (Figure 2.10 A–F): 6S RNA, for example, had two major peaks in fractions 6 and 9 before the manual adjustment and only one in fraction 7 after adjustment. The northern blot data showed a similar, wide peak that was almost fully congruent to the adjusted sedimentation profile of 6S RNA. Similar improvements could be observed for all other tested RNAs, as exemplified by ChiX, *cspE*, CsrB and GcvB. 16S rRNA, on the other hand, already agreed well with the northern blot results before the manual adjustment of the RNA-seq, which, importantly, was still the case after adjustment.

Globally, the manual adjustment smoothened the mentioned distortion around fraction 5 (see 2.4.1), eliminated the "RNA void" of fractions 7 and 8 and generally displayed a changed profile that did not emphasize the 30S subunit peak as strongly anymore (Figure 2.11 A–G). These results show that manual adjustment of the RNA-seq data was necessary in order to accurately reflect the physical distribution of the RNA molecules within the gradient.

2.5 Optimization of the Grad-seq MS protocol

In addition to the RNA-seq protocol, the MS protocol of Grad-seq was optimized in order to improve the detection of proteins and the reconstruction of their migration patterns.

2.5.1 Reduction of samples

Originally, proteins of a Grad-seq experiment were run on an SDS gel, whose lanes were subsequently cut into 10 pieces each (Smirnov et al., 2016). Each of these pieces was then subjected to MS, adding up to >200 MS samples and therefore accounting for a high investment both financially and timewise. Due to the availability of newer MS systems with higher sensitivity, the MS protocol of Grad-seq was changed to in-solution digestion as opposed to digestion from a gel, reducing the number of samples to 22 per gradient (20 fractions, pellet and lysate). A homogenization step using ultrasound followed by centrifugation to remove insoluble material was in-
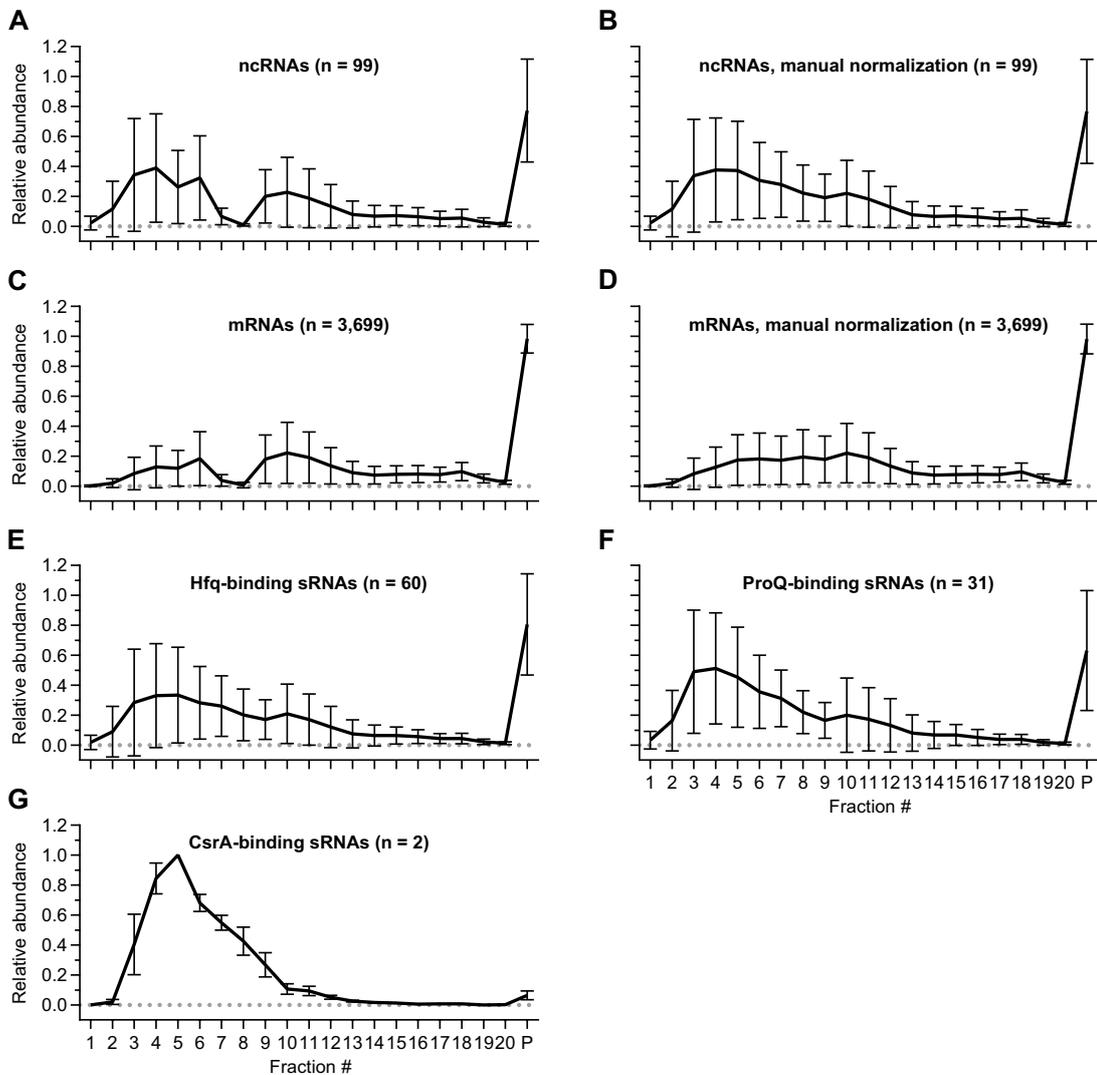
Figure 2.11: Sedimentation profiles of different RNA classes after manual normalization. Comparison of the average sedimentation profiles of confirmed ncRNAs and mRNAs before (**A, C**) and after manual normalization (**B, D**). While mRNAs still show accumulation around the 30S subunit and in the pellet after manual normalization, ncRNAs now show a rather wide spread throughout the gradient with a strong peak around fraction 4 and in the pellet. Average profiles of confirmed Hfq-binding (**E**), ProQ-binding (**F**) and CsrA-binding (**G**) sRNAs after manual normalization reveal smoother profiles when compared to the non-adjusted ones (Figure 2.9 A–C). Profiles are normalized to a maximum of 1. Error bars show SD from the mean.
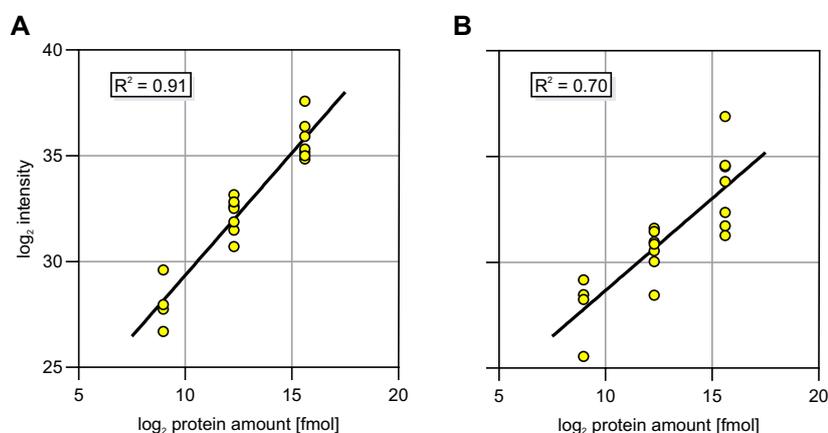
Figure 2.12: Performance test of the UPS2 spike-in. UPS2 spike-in was added to the purified protein of fractions 3 (**A**) and 20 (**B**) of an *E. coli* gradient (compare to Figure 2.4). Following analysis by MS, the measured intensities of the spike-in proteins were plotted against the known amount of the corresponding spike-in proteins. Both tested fractions show excellent linear correlation between the intensities and the known amounts, indicating that the UPS2 spike-in is a suitable tool for the normalization of a Grad-seq experiment. R^2 is the coefficient of determination.

troduced at the same time, which had not been necessary before since the samples were run on a gel.

2.5.2 Use of a complex protein spike-in allows better normalization

Similar to the RNA-seq part of Grad-seq, the MS data has to be normalized to account for variation in sample preparation. Instead of normalizing using the intensities of the corresponding lanes of the Coomassie gel (Smirnov et al., 2016), the commercially available UPS2 spike-in was chosen. It is a complex protein spike-in containing 48 human proteins covering a dynamic range of 5 orders of magnitude, which grants the same advantages as described for the ERCC spike-in (see 2.3.2).

To test the performance of the UPS2 spike-in, 2 fractions with added UPS2 were analyzed by MS. For this, fractions 3 and 20 were chosen since they contained the highest and lowest protein concentrations, respectively, as judged by Coomassie gel analysis (Figure 2.4). 18 out of the 48 spike-in proteins were detected in both cases, covering a dynamic range of 10^3 (Figure 2.12 A and B). The measured intensities further showed a strong linear correlation to the amount of protein added ($R^2 \geq$ 0.70). Thus, the UPS2 spike-in is well-suited to normalize Grad-seq MS datasets,

which is performed by an R-script embedded in GRADitude[2] (Di Giorgio, Hör, Vogel
and Förstner, unpublished).

2.6   Overview of the *E. coli* MS data

2.6.1   Sedimentation of proteins and protein complexes

The MS analysis of the *E. coli* gradient detected 2,145 high confidence proteins, rep-
resenting $\sim$49 % of the proteome as annotated on UniProt (The UniProt Consortium,
2019). The total *E. coli* proteome contains $\sim$56 % and $\sim$31 % soluble and membrane
proteins, respectively (Figure 2.13 A), whereas the Grad-seq proteome exhibited a
bias towards soluble ($\sim$71 %) versus membrane proteins ($\sim$18 %) (Figure 2.13 B). This
enrichment can be explained by the lysate preparation, which removes insoluble par-
ticles such as membranes, their associated proteins and aggregates by centrifugation
prior to gradient centrifugation. In comparison to the RNA-seq data (see 2.4.1), the
MS data already showed rather smooth profiles before normalization based on the
UPS2 spike-in (Figure 2.13 C). The only exception to that were fractions 10 and 12,
which exhibited lower and higher intensities, respectively, than expected based on
the surrounding fractions. After normalization, this distortion was eliminated, how-
ever, and no manual adjustment of the data was necessary (Figure 2.13 D). This
overview of protein intensities further agreed with the abundance estimates from
the gel analysis (Figure 2.4): most proteins sedimented at the top of the gradient,
whereas overall protein abundance dropped toward higher fraction numbers (Figure
2.13 D).

The *E. coli* protein-protein complexome has been studied in great detail for
decades, mostly by low throughput biochemical studies. More recently, however,
global studies trying to analyze all PPIs in a single experiment or series of experi-
ments have become powerful tools to provide overviews of possible interactions (see
1.2.3 and 1.2.4). Together, these studies represent an ever-expanding knowledge of
E. coli protein complexes that is collected in databases such as EcoCyc (Keseler et al.,
2017).

---

[2]https://foerstner-lab.github.io/GRADitude/

A

Localization of the total *E. coli* proteome

2% 1% 3% 3%

6%

10%

49%

26%

- Cytosol (2,368)
- Inner membrane (1,256)
- No information (504)
- Periplasmic space (281)
- Membrane (86)
- Ribosome (59)
- Outer membrane (141)
- Other (161)

Total = 4,856

B

Localization of the *E. coli* Grad-seq proteome

2% 2% 3%

3%

6%

9%

62%

13%

- Cytosol (1,553)
- Inner membrane (321)
- No information (212)
- Periplasmic space (160)
- Membrane (77)
- Ribosome (55)
- Outer membrane (41)
- Other (75)

Total = 2,494

C

Proteins, before UPS2-normalization (n = 2,145)

D

Proteins, after UPS2-normalization (n = 2,145)

E

Figure 2.13: Overview of the MS data. (**A, B**) Comparison of the total and the Grad-seq proteomes reveals enrichment of cytosolic proteins in the Grad-seq experiment. Note that the total number of proteins shown here is different from the number of detected proteins because some proteins have more than one localization assigned. Localization prediction is based on EcoCyc (Keseler et al., 2017). (**C, D**) Comparison of the average sedimentation profile of all proteins before and after UPS2-based normalization, revealing a smoother curve after normalization. Profiles are normalized to a maximum of 1. Error bars show SD from the mean. (**E**) Heat map of selected examples of intact and disassembled complexes as well as the RNA degradosome. Profiles are normalized to a maximum of 1.

This pool of known complexes is well-suited to benchmark the present Grad-seq dataset. For example, ethanolamine ammonia-lyase (EAL) is an enzyme complex composed of 6 EutBC heterodimers and a total molecular weight of ~470 kDa (Akita et al., 2010). In the Grad-seq data, the subunits of EAL were detected with congruent profiles peaking in fraction 5, representing an intact EAL (Figure 2.13 E). Similarly, many other intact complexes were identified. Not all known complexes survived the Grad-seq protocol, however. For example, the DNA polymerase holoenzyme (DNAP) did not show congruent sedimentation of its subunits, indicating that the complex was at least partially disassembled during the experiment. A possible explanation for this is that the DNase treatment of the lysate released DNAP from its substrate, the DNA.

Other cases are more difficult to evaluate. For example, the RNA degradosome consisting of the endoribonuclease RNase E, the RNA helicase RhlB, the 3′→5′ exoribonuclease PNPase and enolase (Bandyra et al., 2013; Mohanty and Kushner, 2018) did not co-migrate within the gradient, which was also observed in *Salmonella* (Smirnov et al., 2016): while RNase E and RhlB sedimented around the 30S subunit, PNPase and enolase sedimented toward the top of the gradient. Intriguingly though, PNPase showed a smaller peak around the 30S subunit too. PNPase is known to exert functions outside of the RNA degradosome, for example, by forming complexes with Hfq (Cameron et al., 2018), which peaked in fraction 4 as well. Enolase is ~20× more abundant than RNase E and RhlB, ~5× more abundant than PNPase (Li et al., 2014) and is, outside of the RNA degradosome, involved in glycolysis (Spring and Wold, 1971). A small increase in enolase intensity around the 30S ribosome could be detected. Indeed, the iBAQ values of RNase E, RhlB, PNPase and enolase, which allow estimation of protein levels (Schwanhäusser et al., 2011), were well within a log-range for the subunit complexes in fraction 12. Overall, this implies that, even though the RNA degradosome seemed disassembled at first glance, all available parameters have to be considered before drawing final conclusions based on Grad-seq data.

To estimate the intactness of complexes more globally, all heterocomplexes, for which all subunits could be detected in the MS data, were tested for co-sedimentation. Of those 107 heterocomplexes, 79 (~74 %) showed high correlation

(Spearman's $\geq$ 0.7), indicating intact complexes (Figure 2.14 A). An important feature of Grad-seq is the size range of complexes that can be analyzed. For example, the succinyl-CoA ligase consisting of SucC and SucD partitioned as a small complex around fraction 3, whereas the >900 kDa FtsH/HflKC metalloprotease complex, in agreement with a previous study (Saikawa et al., 2004), sedimented as a particle of similar size as the 30S subunit, showing the wide range of resolvable complexes (Figure 2.14 B). In agreement with the UV profile and the RNA-seq data, the 30S and 50S subunit proteins peaked around fractions 11 and 16, respectively. These results indicate that the majority of PPIs were kept intact during the preparation of the Grad-seq samples and represent the stable protein complexome of the cell.

2.6.2 Sedimentation of RNA-binding proteins

The vast knowledge on *E. coli* biology also gives the opportunity to specifically investigate molecules with a certain property of interest. RBPs, for example, are major players in many if not most processes within the cell (Babitzke et al., 2019; Holmqvist and Vogel, 2018). To obtain a bird's eye view on possible *in vivo* complex formation of RBPs, the MS data was filtered for proteins with predicted RNA-binding properties based on UniProt (The UniProt Consortium, 2019) and Gene Ontology (Ashburner et al., 2000; The Gene Ontology Consortium, 2019) information (Appendix Figure 7.2). Interestingly, RBPs populated the whole gradient, revealing that some are likely without stable interacting partners, whereas others are involved in complexes of all sizes. This rather naïve approach can already yield functional information: while ribosomal proteins had their peak abundances in the pellet fraction, proteins involved in ribosome maturation were not found in the pellet, showing that the latter are not involved in actively translating ribosomal complexes but rather dissociate after fulfilling their respective functions.
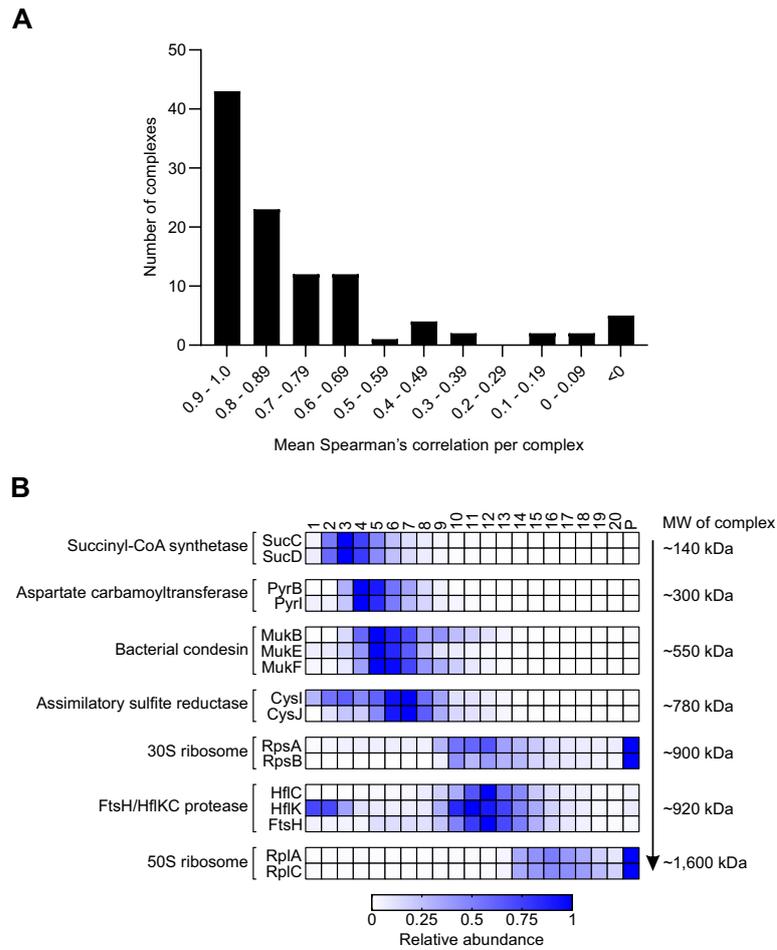
Figure 2.14: Determination of the intactness of complexes. (**A**) Spearman correlation analysis was performed for all heterocomplexes listed in EcoCyc (Keseler et al., 2017), for which all subunits were detected in Grad-seq. The majority of these complexes show high correlation of ≥0.7 and only few show anti-correlation (r < 0), indicating that most detected complexes survived the Grad-seq protocol. (**B**) Heat map of selected intact complexes spanning a wide range of molecular weights (MW). Profiles are normalized to a maximum of 1.
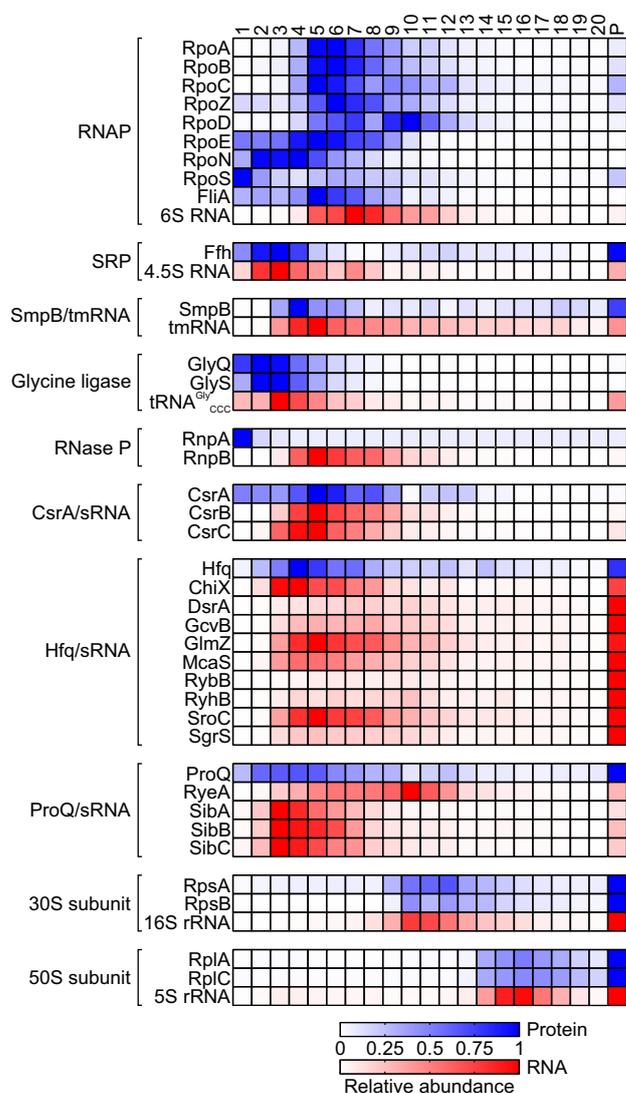
Figure 2.15: Overview of conserved RNPs. Heat map showing major stable RNPs of *E. coli*. Except for RNase P and some of the σ-factors of RNA polymerase (RNAP), all RNPs show strong correlation between the corresponding protein and RNA subunits. It is of note that both Hfq and ProQ as well as their associated sRNAs show high abundance in the pellet, indicating potential ribosome association. SRP, signal recognition particle. Profiles are normalized to a maximum of 1.

2.7 Combined analysis of RNA-seq and MS data

2.7.1 Overview of conserved RNPs

The possibility of analyzing complex formation of RNAs is one of the main benefits of Grad-seq compared to other methods (see 1.2.5)(Smirnov et al., 2017a). Similar to the PPIs described in 2.6.1, the major RNPs of *E. coli* were analyzed with regards to their gradient sedimentation. As observed in the laboratory before (Figure 2.5 A and C), the RNAP consisting of the α-, β-, β'- and ω-subunits (RpoA, RpoB, RpoC and RpoZ, respectively) co-migrated with 6S RNA (Figure 2.15), which is a negative regulator of transcription that functions by competing for promoter binding of RNAP-σ^{70} (Wassarman, 2018). It is of note, however, that σ^{70} (RpoD) showed a second peak around fraction 10, outside of the RNAP, which was not detected by western blotting (Figure 2.5 C). Detailed analysis of the raw MS data confirmed that the measured peptides indeed derived from RpoD with high confidence. A similar, but weaker, secondary peak was also observed for RpoC (Figure 2.15), yet it would be speculation to interpret this to be a functional form of RpoD without any further experiments.

Interestingly, while the σ-factors RpoD, σ^{24} (RpoE) and σ^{28} (FliA) were detected to co-migrate with RNAP, this was not the case for the majority of the signal of σ^{54} (RpoN) or σ^{S} (RpoS). Finally, the minor σ-factor σ^{19} (FecI), which is involved in the uptake of ferric citrate (Braun et al., 2003), was not detected in the dataset. The ribosomal subunits, the signal recognition particle (SRP) as well as the SmpB-tmRNA and glycine-tRNA ligase-tRNA^{Gly} RNPs are other major RNPs whose subunits were found to co-sediment within the gradient. In contrast, RnpA (the protein factor of RNase P) was only detected at very low intensity in the first 3 fractions, making an estimation of its binding to RnpB (the RNA part of RNase P) difficult. Similarly, RnpA could not be detected in *Salmonella* (Smirnov et al., 2016).

The three major known regulatory RBPs of *E. coli*, CsrA, Hfq and ProQ, are known to bind specific subsets of sRNAs (Holmqvist et al., 2018; Melamed et al., 2020; Potts et al., 2017; Tree et al., 2014). CsrA showed a rather wide peak and no abundance in the pellet fraction. Its main antagonists, the sRNAs CsrB and CsrC, exhibited a narrower distribution in the gradient with a peak in fraction 5—the

main peak of CsrA—and no abundance in the pellet, indicating a stable RNP. The wider peak of CsrA might be explained by binding to mRNAs of all lengths, causing CsrA-RNA complexes of a wide range of sizes (Romeo and Babitzke, 2018).

Hfq binds more sRNAs than CsrA and is generally thought to be necessary for the majority of sRNA-based regulation within *E. coli* (Hör et al., 2020b; Kavita et al., 2018; Updegrove et al., 2016). In the present Grad-seq data, Hfq showed peaks in fraction 4 and the pellet. Interestingly, the ribosome association indicated by the accumulation in the pellet had already been described even before the function of Hfq was known (Kajitani et al., 1994). On average, this sedimentation was mirrored by the known Hfq-binding sRNAs (Figure 2.11 E). The individual sRNAs, however, exhibited sedimentation profiles that strongly differed between each other (Figure 2.15). For example, the abundant ChiX, which is possibly the strongest Hfq binder (Małecka et al., 2015), showed its peak abundance in fraction 4, with lower levels detected in the pellet as well. RybB, on the other hand, was almost exclusively present in the pellet fraction, which indicated that it preferentially bound to the ribosome-associated subpopulation of Hfq.

In comparison to Hfq, a lot less is known about the function of ProQ and its associated sRNAs. The best understood ProQ-binding sRNAs are probably the wealth of antisense sRNAs functioning as the antitoxin in type I toxin-antitoxin systems (Holmqvist et al., 2018; Melamed et al., 2020; Smirnov et al., 2016). In the *E. coli* Grad-seq data, ProQ-binding sRNAs showed a higher average abundance toward the top of the gradient around fraction 4, which is also where ProQ was found to peak (Figures 2.11 F and 2.15). The ProQ-binding antitoxins SibA, SibB and SibC (Harms et al., 2018) were found in this region of the gradient as well. In contrast, RyeA, which is the antitoxin of SdsR (Choi et al., 2018), was shown to bind ProQ (Holmqvist et al., 2018), but sedimented away from it in the gradient. Interestingly, similar to Hfq, high levels of ProQ were observed in the pellet fraction, which might indicate ribosome-association and which could be verified by polysome gradient fractionation (Appendix Figure 7.3).
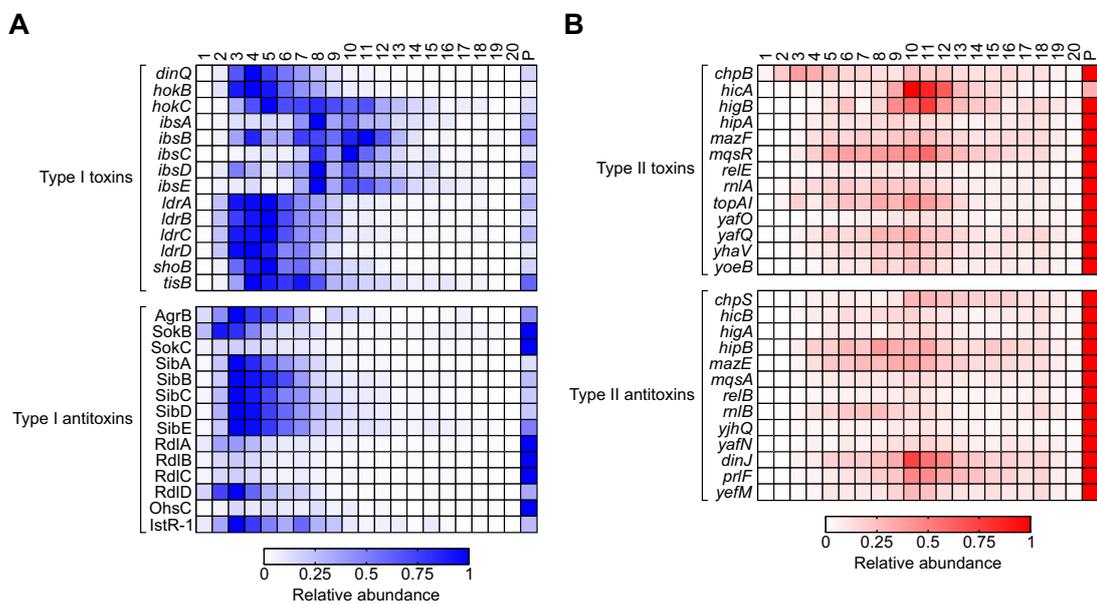
Figure 2.16: Sedimentation of toxin/antitoxin RNAs. (A) Heat map showing the sedimentation profiles of type I toxin mRNAs (top) and their corresponding antitoxin antisense RNAs (bottom). The mRNAs do not accumulate in the pellet of the gradient, indicating that they are not actively translated. **(B)** Heat map showing the sedimentation profiles of type II toxin mRNAs (top) and their corresponding antitoxin mRNAs (bottom). In this case, the antitoxins are proteins that need to be translated in order to inactivate the toxin. Therefore, both the toxin mRNAs and the antitoxin mRNAs accumulate in the pellet of the gradient, indicating active translation. Profiles are normalized to a maximum of 1.

2.7.2 mRNAs with peculiar gradient sedimentation

To take advantage of the information about RNA sedimentation, the list of detected mRNAs was filtered according to their peak abundance. For translated mRNAs, a peak is expected in the pellet fraction because of the association with ribosomes. 170 (~4.6 %) mRNAs were found to have a relative abundance of <1 in the pellet fraction, indicating that they might not be efficiently translated or possibly degraded, moving them away from the pellet fraction. Interestingly, 15 of the 16 detected type I toxin/antitoxin (TA) system toxin mRNAs were among these 170 mRNAs (Figure 2.16 A). Type I TA systems consist of a toxic protein, whose expression is repressed by a noncoding antisense RNA (the antitoxin) that blocks translation of the toxin (Berghoff and Wagner, 2017; Gerdes and Wagner, 2007; Harms et al., 2018). In the Grad-seq data, the majority of the antitoxin RNAs sedimented together with their respective toxin mRNAs (Figure 2.16 A), suggesting formation of translationally inactive complexes. These complexes, however, are rapidly cleaved by RNase III (Berghoff and Wagner, 2017; Gerdes and Wagner, 2007), making it unlikely they would survive the Grad-seq protocol. As mentioned before, ProQ is known to bind many of the antitoxin RNAs (Holmqvist and Vogel, 2018; Melamed et al., 2020; Smirnov et al., 2016), which explains why some of the antitoxins sedimented around fraction 4 (compare to Figure 2.15). The only type I toxin mRNA that actually peaked in the pellet fraction was *hokD*, which, however, does not possess an antitoxin RNA (putative "SokD") on the opposite strand (Pedersen and Gerdes, 1999). In contrast to type I TA systems, both the toxins and the antitoxins of type II TA systems are proteins (Harms et al., 2018). Consequentially, both mRNA partners of type II TA systems were found to have their peak abundance in the pellet fraction (Figure 2.16 B).

2.8 RyeG encodes a toxic, prophage-derived small protein

Since the short list of sRNAs is easier to survey than the list of mRNAs, manual inspection of sRNAs with unexpected (compare to Figure 2.11 B) sedimentation profiles was carried out. One of the sRNAs with a particularly interesting gradient sedimentation was RyeG. It only co-migrated with the 30S subunit and additionally

A

B
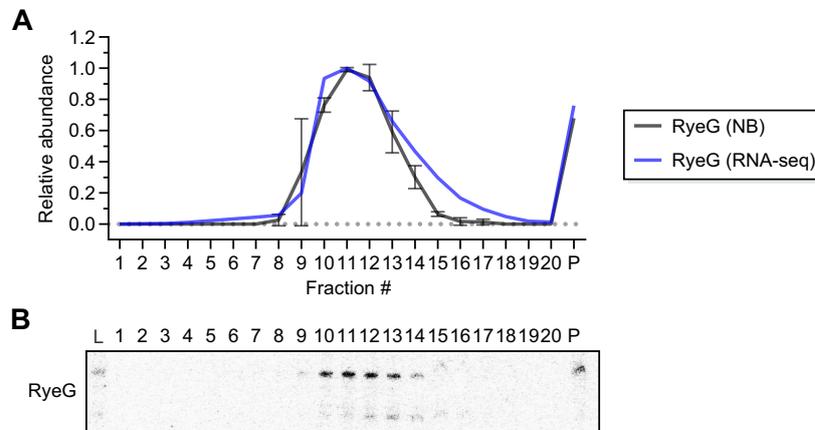
Figure 2.17: RyeG co-migrates with the 30S subunit. (**A**) Sedimentation profiles of RyeG according to northern blot (NB, gray, n = 2) and RNA-seq (blue) analyses show co-migration of RyeG with the 30S subunit around fraction 11. Profiles are normalized to a maximum of 1. Error bars show SD from the mean. (**B**) NB analysis confirming the results from the sequencing (quantification shown in (**A**)).

showed slight abundance in the pellet (Figure 2.17 A), which, importantly, could be validated by northern blotting (Figure 2.17 A and B).

2.8.1   RyeG is a prophage-encoded RNA

The *ryeG* gene is located on the antisense strand between *yfdI* and *tfaS* in the cryptic prophage CPS-53 (Figure 2.18 A), which is only present in the genomes of K-12 strains. CPS-53 seems to be in the process of degradation since many of its genes are pseudogenes and it further is defective for lysis (Panis et al., 2007). Still, it provides limited functions, such as increasing H_2O_2 and acid resistance (Wang et al., 2010) or encoding genes able to inhibit initiation of chromosomal replication when overexpressed (Noguchi and Katayama, 2016).

RyeG itself was first reported as "IS118" in a bioinformatics study predicting sRNAs in *E. coli* (Chen et al., 2002) and only one study investigated its function by overexpression, showing decreased biofilm formation and motility (Bak et al., 2015). The 199 nt long RNA, which is fully conserved among CPS-53-carrying K-12 strains, is under the control of a σ^{70} promoter with an extended –10 box close to the consensus of TGNTATAAT (Burr et al., 2000; Mitchell et al., 2003) but without an obvious –35 box (Figure 2.18 B). Its annotated transcriptional start site (TSS) could be confirmed by differential RNA-seq (Thomason et al., 2015). Interestingly, RyeG
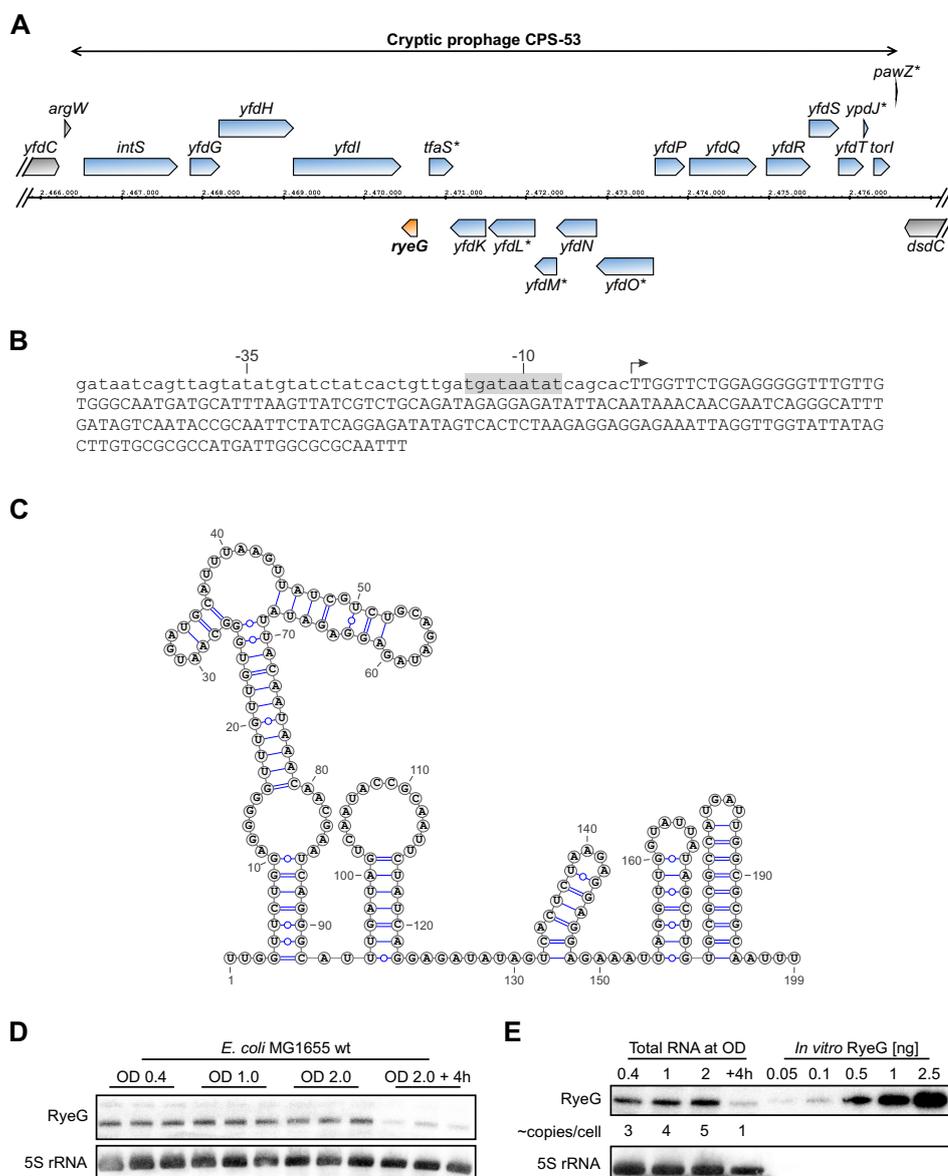
Figure 2.18: RyeG is a prophage-encoded RNA. (**A**) Locus of the cryptic prophage CPS-53 in the genome of *E. coli* K-12 MG1655. *ryeG* (orange) is encoded on the antisense strand in between *yfdI* and *tfaS*. Genes outside of CPS-53 are shown in gray, genes within CPS-53 are shown in blue. Asterisks denote pseudogenes. (**B**) Locus of *ryeG*. The arrow indicates the transcriptional start site (+1) of *ryeG* and is followed by the *ryeG* sequence in capitals. The upstream sequence of the +1 site (smaller case letters) includes an extended –10 box (highlighted in gray) and no –35 box. (**C**) Predicted secondary structure of RyeG based on RNAfold (Lorenz et al., 2011), illustrated with VARNA (Darty et al., 2009). RyeG folds into a tight structure with several stem-loops including a ρ-independent terminator. (**D**) Northern blot analysis of the expression of RyeG over time in wild-type (wt) *E. coli* MG1655 reveals constant expression while the cells are still growing ($OD_{600\,nm}$ of 0.4–2.0), whereas its expression is downregulated in late stationary phase ($OD_{600\,nm}$ of 2.0 + 4 h). 5S rRNA served as loading control. (**E**) To estimate the RyeG copy numbers per cell, total RNA of a growing culture was compared to defined amounts of *in vitro*-transcribed RyeG by northern blotting. This revealed low numbers of approximately 1–5 copies/cell, depending on the growth phase. 5S rRNA served as loading control.

is predicted to fold into a highly structured RNA (Figure 2.18 C) with several stem loops, including a ρ-independent terminator. Reminiscent of highly structured ProQ-binding sRNAs (Smirnov et al., 2016), RyeG was indeed found to associate with ProQ (Holmqvist et al., 2018), but also with Hfq (Michaux, Hör and Vogel, unpublished).

In agreement with the predicted transcriptional activation by σ^{70}, RyeG could be detected in similar quantities throughout growth in rich medium, whereas it showed strongly reduced expression in late stationary phase (Figure 2.18 D). To quantify the cellular levels of RyeG over growth, its expression was compared to *in vitro*-synthesized RyeG (Figure 2.18 E), revealing low numbers of ∼3–5 copies per cell during growth and ∼1 copy in late stationary phase.

2.8.2   RyeG associates with 30S subunits and 70S ribosomes

To further investigate the migration behavior of RyeG, the Grad-seq MS data were searched for proteins deriving from CPS-53 that might explain the observation. With IntS, YfdH and YfdI, three CPS-53-derived proteins could be detected. Surprisingly, IntS, which is the integrase of CPS-53 (Panis et al., 2010b), co-migrated with RyeG and the 30S subunit (Figure 2.19 A). The sedimentation of IntS, however, could not be confirmed by glycerol sedimentation of a lysate obtained from an *intS*-3xFLAG strain (Figure 2.19 B). Therefore, the potential association of RyeG with the 30S subunit was studied in more detail by heterologously expressing RyeG from a plasmid in *E. coli* 536, a uropathogenic strain that does not contain the CPS-53 prophage. Northern blotting revealed that RyeG also co-migrated with the 30S subunit in *E. coli* 536, which eliminated IntS as a potential interactor (Figure 2.19 C).

To experimentally validate binding of RyeG to the ribosome, *in vitro* reconstitution was performed using purified ribosomes from a Δ*ryeG* strain and *in vitro*-transcribed RyeG. The reconstituted complexes were subsequently run on a polysome gradient to analyze the binding (Figure 2.19 D). Even though the purified ribosomes contained mostly 70S ribosomes and little 30S subunits, a clear enrichment of RyeG at the 30S subunit could be detected. However, the majority of the RyeG signal was detected together with the 70S ribosomes, which might be explained either by binding to the 30S subunit of the 70S ribosome or by formation of translating complexes. It is important to note that the purification of ribosomes was
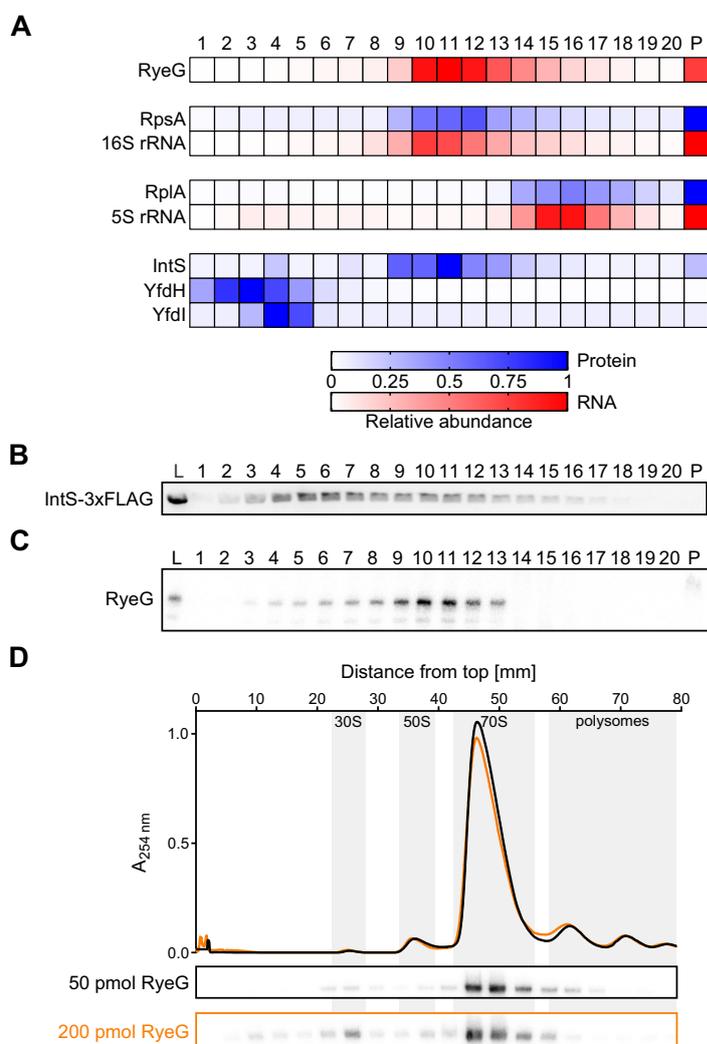
Figure 2.19: Verification of the RyeG-30S subunit association. (**A**) Heat map showing the sedimentation profiles of RyeG, the 30S subunit (represented by RpsA and 16S rRNA), the 50S subunit (represented by RplA and 5S rRNA) and all detected proteins of CPS-53 (IntS, YfdH and YfdI). RyeG and IntS co-migrate with the 30S subunit. Profiles are normalized to a maximum of 1. (**B**) Glycerol gradient analysis followed by western blotting of an *intS*-3xFLAG strain shows sedimentation of IntS-3xFLAG around fraction 5, disagreeing with the MS results shown in (**A**). (**C**) Glycerol gradient analysis followed by northern blotting of heterologous expression of RyeG in *E. coli* 536, which does not carry the CPS-53 prophage, verifies the 30S subunit association of RyeG independent of other phage proteins. (**D**) *In vitro* reconstitution of 50 (black) or 200 pmol (orange) *in vitro*-transcribed RyeG and 200 pmol of purified ribosomes obtained from a Δ*ryeG* strain followed by sucrose polysome gradient analysis reveals binding of RyeG to the 30S subunit and 70S ribosomes. L, lysate (input control). P, pellet.

performed using sucrose cushions that should not let smaller molecules than ribosomal subunits pass through, leaving tRNAs in the supernatant (Mehta et al., 2012). Consequently, formation of translating complexes seemed unlikely. These results demonstrate that RyeG is able to directly bind to 30S subunits and 70S ribosomes.

2.8.3   Overexpression of RyeG prolongs lag phase

To identify potential functions of RyeG, growth curve experiments were performed comparing constitutive overexpression of RyeG from a high copy plasmid against the same strain carrying a control plasmid (Figure 2.20 A). This overexpression resulted in a strongly increased lag phase and was also true for *E. coli* 536 and BL21(DE3) as well as *S.* Typhimurium (Figure 2.20 B–D), all of which do not encode for CPS-53. Lag phase is the phase of bacterial growth that prepares the cells for exponential growth after being transferred to new medium (Bertrand, 2019). During lag phase, cells do not divide but are metabolically active and grow in size. Increased lag times can be the result of several underlying causes such as stress or injury of the cells (Bertrand, 2019), which might be caused by overexpression of RyeG. To test whether RyeG prolongs lag time by slowing down the growth of cells before the first division, a refreshed culture was grown in rich medium for 1.5 h before visualization by microscopy (Figure 2.20 E and F). Indeed, cells appeared smaller and less dividing cells were found in the overexpression strain compared to the control. These results indicate a toxic effect of RyeG.

2.8.4   RyeG contains ORFs and can be translated

The ability of RyeG to bind to 30S subunits and 70S ribosomes indicated that it might either contain an ORF or that it can bind and potentially modulate the ribosome in another way. To assess the first option, ORFfinder[3] was used to predict all possible ORFs within RyeG. This search returned five different ORFs (Appendix Figure 7.4 A–E), of which ORF1 could be excluded immediately since the software assigns this ORF in a way that it might start upstream of the +1 site of the input sequence. ORF2 has a GUG start codon at position +23 of RyeG and spans almost the entire

---

[3]https://www.ncbi.nlm.nih.gov/orffinder/
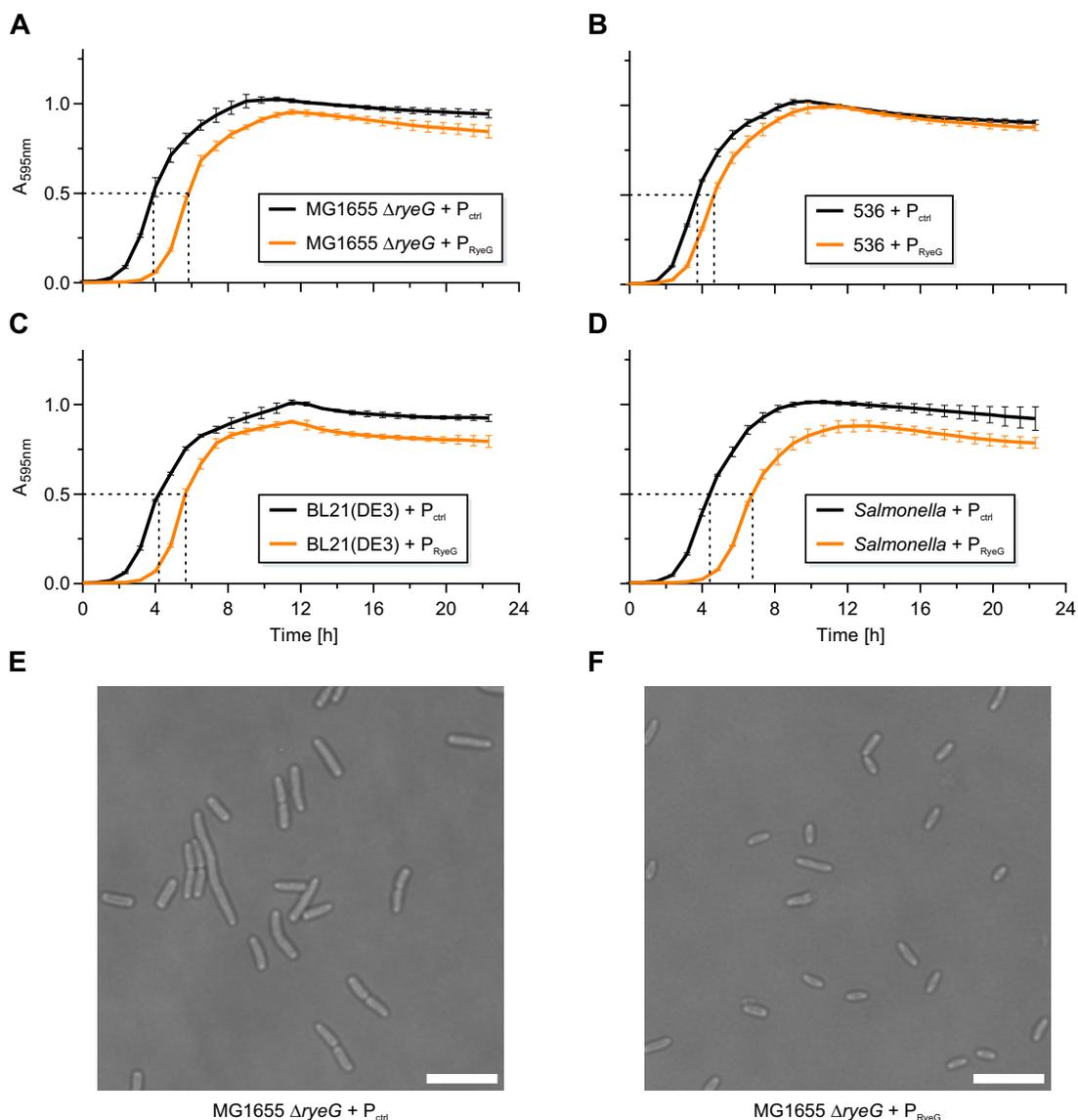
Figure 2.20: Overexpression of RyeG prolongs lag phase. Growth curves in rich medium of *E. coli* MG1655 Δ*ryeG* (**A**), *E. coli* 536 (**B**), *E. coli* BL21(DE3) (**C**) and *Salmonella* Typhimurium (**D**) carrying a control plasmid (black) or a RyeG overexpression plasmid (orange), measured in a plate reader. Overexpression of RyeG strongly prolongs the lag phase of these bacteria, as indicated by delayed reaching of mid-exponential growth (dashed lines). As soon as the bacteria overcome the lag phase, growth is comparable between the controls and the overexpression strains. Error bars show SD from the mean, n = 3. (**E, F**) Representative microscopy images of *E. coli* MG1655 carrying a control or RyeG overexpression plasmid 1.5 h after refreshing an overnight culture in fresh medium. The overexpression strain shows smaller cells and less division events, agreeing with the increased lag phase observed in (**A**). Scale bars represent 7.5 μm.

A

B

```
                                        -14
                                         |
UUGGUUCUGGAGGGGGUUUGUUGUGGGCAAUGAUGCAUUU
AAGUUAUCGUCUGCAGAUAGAGGAGAUAUUACAAUAAACA
ACGAAUCAGGGCAUUUGAUAGUCAAUACCGCAAUUCUAUC
AGGAGAUAUAGUCACUCUAAGAGGAGGAGAAAUUAGGUUG
GUAUUAUAGCUUGUGCGCGCCAUGAUUGGCGCGCAAUUU
```
ORF2

C

```
                                        -16
                                         |
UUGGUUCUGGAGGGGGUUUGUUGUGGGGCAAUGAUGCAUUU
AAGUUAUCGUCUGCAGAUAGAGGAGAUAUUACAAUAAACA
ACGAAUCAGGGCAUUUGAUAGUCAAUACCGCAAUUCUAUC
AGGAGAUAUAGUCACUCUAAGAGGAGGAGAAAUUAGGUUG
GUAUUAUAGCUUGUGCGCGCCAUGAUUGGCGCGCAAUUU
```
ORF3

Figure 2.21: RyeG can be translated. (**A**) Toeprinting assay of *in vitro*-transcribed RyeG. In presence of initiator tRNA and 30S subunits, a stop in the reverse transcription at an A nucleotide at position +37 of RyeG is observed, indicating formation of an initiation complex. (**B**) Sequence of RyeG with ORF2 (Appendix Figure 7.4 B) highlighted in bold. The putative ribosome-binding site is highlighted in gray. The A nucleotide identified in (**A**) is highlighted in blue, showing that the first nucleotide of the start codon of ORF2 is located 14 nt upstream of it. (**C**) Sequence of RyeG with ORF3 (Appendix Figure 7.4 C) highlighted in bold. The putative ribosome-binding site is highlighted in gray. The A nucleotide identified in (**A**) is highlighted in blue, showing that the first nucleotide of the start codon of ORF3 is located 16 nt upstream of it.

RNA. Importantly, a potential ribosome-binding site (RBS) is located 6 nt upstream of the start codon, which additionally seems accessible based on secondary structure prediction (Figure 2.18 C). The putative protein encoded by ORF2 is 48 aa in length. ORF3 starts only 2 nucleotides upstream of ORF2 and contains a UUG start codon. A second UUG codon is present directly upstream, which, however, is an unlikely start codon given the short distance of 4 nt to the predicted RBS. This RBS is predicted to be less accessible than the RBS of ORF2 (Figure 2.18 C) and the ORF codes for a 19 aa protein. ORF4 and ORF5 do not contain obvious RBSs in the preferred distance of ∼3–7 nt (Chen et al., 1994) upstream of the start codon. While an RBS is not strictly necessary to allow translation (Omotajo et al., 2015), ORF4 and ORF5 appeared to be worse candidates than ORF2 and ORF3.

To analyze whether translation initiation can take place on RyeG, toeprinting assays were performed using *in vitro*-synthesized RyeG and purified 30S subunits. If initiation can occur, reverse transcription of an mRNA in presence of 30S and initiator tRNA fMet-tRNA$_i^{Met}$ leads to a stop at a distance of 14–16 nt downstream of the +1 site (Hartz et al., 1988), which can be used to identify the start codon.

A

SD-mut
GAAAA
↑↑↑↑↑
UUGGUUCUGGAGGGGGUUUGUU...
ORF2
ORF3

ORF2-stop
A
↑
GUG GGC AAU GAU GCA UUU AAG UUA...

ORF3-stop
A
↑
UUG UGG GCA AUG AUG CAU UUA AGU UAU...

B
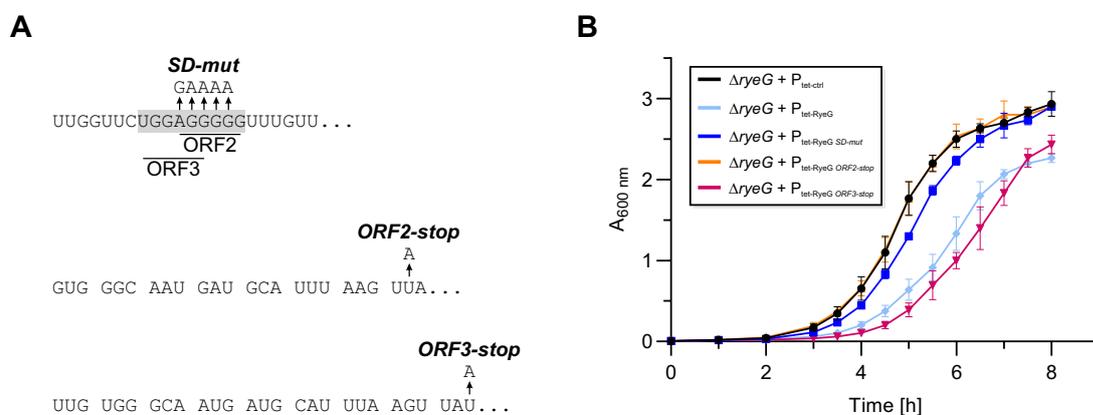
Figure 2.22: ORF2 is solely responsible for the phenotype of RyeG. (**A**) To analyze whether ORF2, ORF3 or neither of them is responsible for the phenotype of RyeG overexpression (Figure 2.20), several mutants were constructed on tetracycline-inducible plasmids. *SD-mut* disrupts the overlapping putative ribosome-binding sites of ORF2 and ORF3 without disrupting the bulge structure they are in (compare to Figure 2.18 C). *ORF2-stop* and *ORF3-stop* introduce premature stop codons in ORF2 and ORF3, respectively, by exchanging single nucleotides. (**B**) Growth curves of RyeG variants overexpressed from tetracycline-inducible plasmids in *E. coli* MG1655 Δ*ryeG*. Induction was started by addition of 200 ng/ml doxycycline to the medium in which the growth curves were performed. The results show that the increase in lag phase can only be observed when ORF2 is translatable. Error bars show SD from the mean, n = 3.

Toeprinting of RyeG revealed a strong toeprint at position +37 in presence of 30S and fMet-tRNA$_i^{Met}$ but not without the initiator tRNA, indicating assembly of an initiation complex (Figure 2.21 A). The A nucleotide at position +37 is located 14 nt and 16 nt downstream of the start codons of ORF2 (Figure 2.21 B) and ORF3 (Figure 2.21 C), respectively, suggesting that both ORFs can potentially be translated.

Since the toeprinting assay did not unambiguously reveal which of the ORFs is the ORF being translated, three different mutant versions of RyeG were cloned into a tetracycline-inducible plasmid, which had to be used due to the inability of cloning certain RyeG versions into a constitutive overexpression plasmid. *SD-mut* is a mutant that changes 5 nt in the shared RBS of ORF2 and ORF3 without disruption of the predicted secondary structure (Figure 2.22 A). If the observed toxicity of RyeG is dependent on the translation of a small protein, this mutation should render RyeG non-toxic. Similarly, *ORF2-stop* and *ORF3-stop* contain a single nucleotide exchange introducing a premature stop codon in ORF2 and ORF3, respectively. If one of the ORFs is responsible for the toxicity of RyeG, the corresponding stop mutant should render the plasmid non-toxic. As observed before, wild-type RyeG caused a strong

increase in lag time when expressed from the tetracycline-inducible plasmid (Figure 2.22 B).

This toxicity was mostly remedied by *SD-mut*, which grew very similar to the control, even though a slight increase in lag time remained. Interestingly, *ORF2-stop* had an even stronger effect and grew without any obvious defect when compared to the control. In stark contrast to this, *ORF3-stop* did not decrease the toxicity of RyeG but instead even slightly increased the lag time compared to wild-type RyeG. Taken together, RyeG encodes for a 48 aa small protein encoded by an ORF here denoted as ORF2. Expression of this protein is toxic to the cell and increases lag time in rich medium. Importantly, all tries to clone or transform inducible or constitutive protein overexpression plasmids containing ORF2 failed. The strong RBS present on these plasmids in combination with the overexpression likely led to levels of the protein that were too high to obtain transformants.

2.9   YggL is a conserved ribosome-binding protein

As shown in 2.6.1, Grad-seq allows the analysis of intact soluble complexes. It is therefore possible to predict whether a protein is likely involved in a complex, just based on its positioning within the gradient—the "guilt-by-association" logic. For example, a <20 kDa protein should sediment around <3S given a slightly elongated shape, meaning it should sediment at the top of the gradient (Erickson, 2009). Conversely, this implies that a <20 kDa protein, whose peak is detected in higher fraction numbers, is likely involved in a complex.

2.9.1   Small proteins with unexpected gradient sedimentation

Thus, to predict possible new complexes, the MS data were filtered to only include proteins <20 kDa that showed their peaks in fractions 4 and above, resulting in a list of 97 proteins (Figure 2.23). Unsurprisingly, 42 of the 97 small proteins were ribosomal proteins with 4 more being known to be ribosome-associated: Hsp15 (HslR), RsfS, the L31 paralog YkgM and Rmf. Hsp15 and RsfS were co-sedimenting with the 50S subunit, which was also shown in earlier studies (Häuser et al., 2012; Jiang et al., 2006, 2009; Korber et al., 2000). YkgM could only be detected in fraction 15
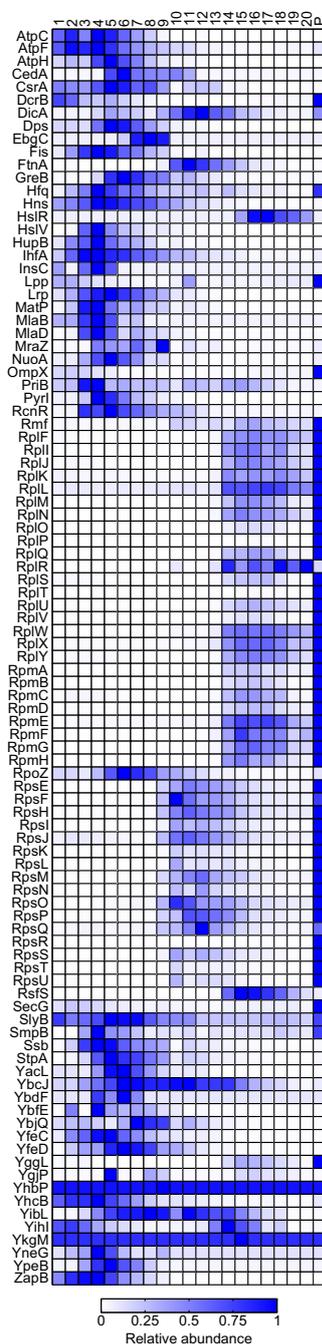
Figure 2.23: Sedimentation profiles of fast-sedimenting small proteins. Heat map showing the sedimentation profiles of 97 proteins with a molecular weight <20 kDa and peaks in fraction 4 or higher. Many ribosomal proteins (Rpl*, Rpm* and Rps*) are among these proteins as well as several proteins of unknown function, for example, YggL. All of these proteins are predicted to form complexes. Profiles are normalized to a maximum of 1.

and therefore at the 50S subunit. Considering its probable function as an alternative L31 protein (Hensley et al., 2012), this might be a true interaction rather than a detection by chance. In contrast, Rmf showed a clear peak in the pellet fraction, thereby matching its function in the formation of 100S ribosomes (inactive dimers of 70S ribosomes) (Ueta et al., 2005, 2008). Other expected proteins included the RNAP-interacting proteins RpoZ (the ω-subunit of RNAP) (Mathew and Chatterji, 2006), GreB (a transcription elongation factor) (Opalka et al., 2003) and CedA (a regulator of cell division) (Abe et al., 2016), all of which co-sedimented with the RNAP core enzyme (compare to Figure 2.15).

However, many small proteins of unknown function were also found to sediment in higher fraction numbers of the gradient, implying their involvement in complexes: YacL, YbdF and YfeD showed similar sedimentation and co-migrated with the RNAP (Figure 2.23). Others, such as YbcJ and YibL were detected with high intensities throughout the gradient (average $\log_{10}$ intensity of $\sim$8.3), showing very broad distributions without clear peaks, indicating they could be involved in several different complexes or that they formed artificial aggregates. For YbcJ, several interacting partners like the degradosome or the 50S subunit have been described (Jiang et al., 2007; Hu et al., 2009) which could explain the observations. In contrast, YibL was reported to co-migrate with the 50S subunit in sucrose gradients (Jiang et al., 2006), contrasting the sedimentation shown here.

2.9.2   YggL binds the 50S subunit

Another one of the identified proteins was YggL. YggL is a 108 aa ($\sim$13 kDa) protein that co-migrated with the 50S subunit and had strong abundance in the pellet as well, mirroring the sedimentation of other 50S components (Figures 2.23 and 2.24 A). To verify this observation, a glycerol gradient was run with a lysate obtained from a *yggL*-3xFLAG strain followed by western blotting (Figure 2.24 B). Interestingly, the profile of YggL-3xFLAG supported potential 50S subunit association, while the abundance in the pellet fraction was not as high as in the MS data. YggL was suggested to be involved in late 50S subunit assembly or final maturation of the 70S ribosome based on its localization in 50S and 70S fractions in a sucrose gradient (Chen and Williamson, 2013), which agrees with the data obtained here. Of note,

A

B
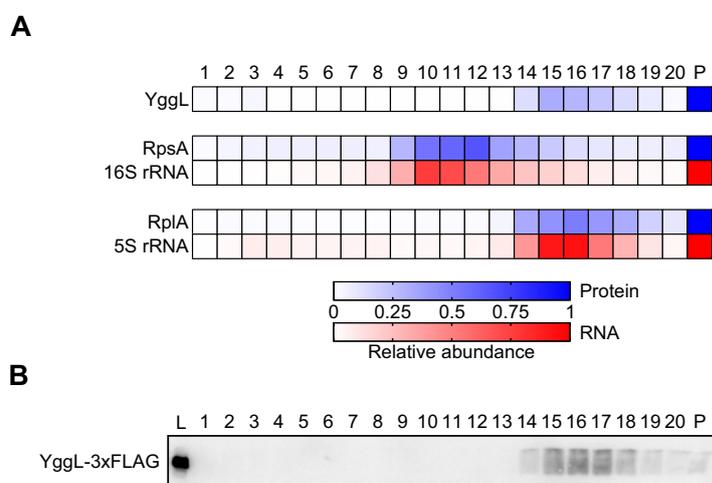
Figure 2.24: YggL co-migrates with the 50S subunit. (**A**) Heat map showing the sedimentation profiles of YggL, the 30S subunit (represented by RpsA and 16S rRNA) and the 50S subunit (represented by RplA and 5S rRNA). YggL co-migrates with the 50S subunit. Profiles are normalized to a maximum of 1. (**B**) Glycerol gradient analysis followed by western blotting of a *yggL*-3xFLAG strain shows sedimentation of YggL-3xFLAG around fraction 16, where the 50S subunit sediments as well. L, lysate (input control). P, pellet.

YggL-3xFLAG shows a peculiar smear in western blots of gradient fractions (not visible in the lysate control), which could indicate phosphorylation or degradation of the protein. Analysis of the phosphorylation state by phosphatase assays led to degradation of YggL and could therefore not confirm or disprove phosphorylation of YggL.

yggL is encoded on the antisense strand in between *yggN* (a protein of unknown function) and *trmI* (a tRNA methyltransferase) (Figure 2.25 A). Differential RNA-seq (dRNA-seq) (Thomason et al., 2015) revealed 3 potential TSSs and 1 potential cleavage site upstream of *yggL* (Figure 2.25 A and B). Of the 3 TSSs, TSS1 and TSS2 are upstream of *trmI*, but only transcription from TSS2 includes *yggL*, generating a bicistronic *trmI-yggL* mRNA. TSS3 is internal to the *trmI* coding sequence (CDS), whereas the cleavage site is $\sim$70 nt upstream of the *yggL* start codon. TSS2 and TSS3 both possess predicted σ^{70}-dependent –10 and –35 boxes (Mazumder and Kapanidis, 2019) (Figure 2.25 B). Interestingly, transcription from TSS2 seems to be stronger at an $OD_{600\,nm}$ of 0.4 (exponential growth) than at an $OD_{600\,nm}$ of 2.0 (early stationary phase), while transcription from TSS3 seems to be similar at both points during growth (Thomason et al., 2015). Moreover, transcription from TSS2 at an $OD_{600\,nm}$
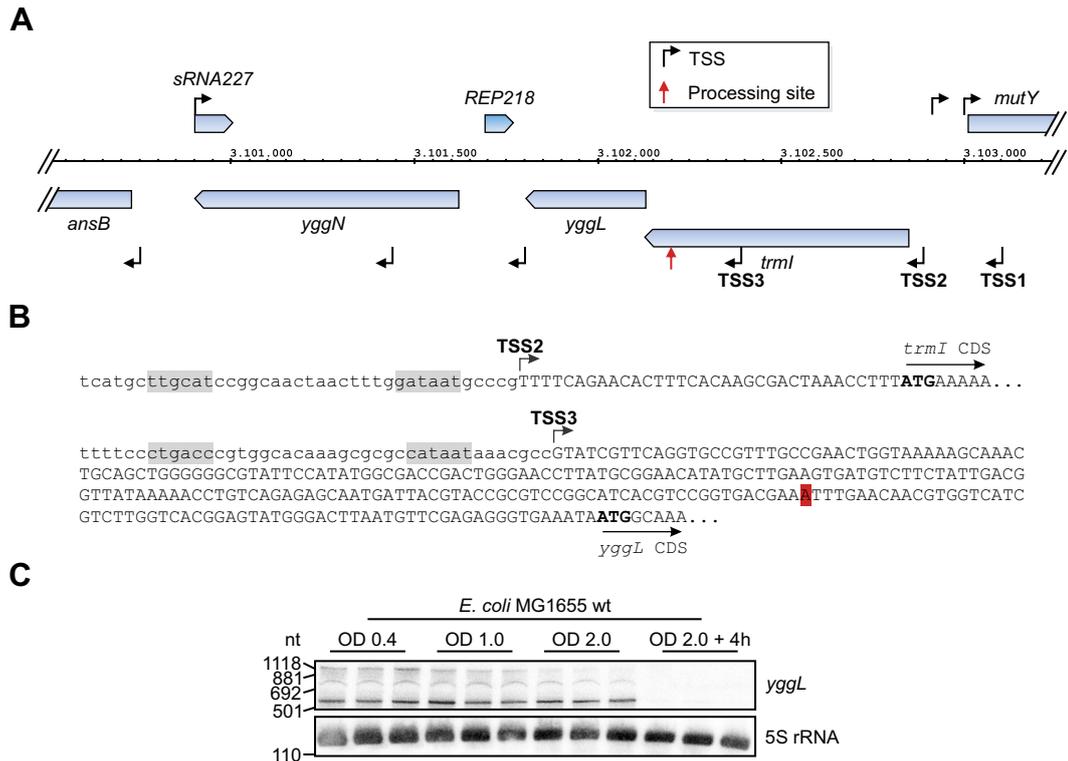
Figure 2.25: Locus and expression of *yggL*. (A) Locus of *yggL* in the *E. coli* MG1655 chromosome. *yggL* is encoded on the antisense strand between *yggN* and *trmI*. There are three transcriptional start sites (TSSs) upstream of *yggL* (Thomason et al., 2015), of which TSS2 and TSS3 include *yggL*. Additionally, a processing site downstream of TSS3 leads to a shortened transcript that includes *yggL* and which can originate from transcripts of either TSS2 or TSS3. **(B)** Locus upstream of *yggL*. TSS2 has a predicted σ^{70}-binding site (highlighted in gray) and transcribes the *trmI-yggL* mRNA (in capitals). Likewise, TSS3 has a predicted σ^{70}-binding site (highlighted in gray) and transcribes *yggL* with an extended 5′ UTR, that can be shortened by cleavage at an A nucleotide (highlighted in red). **(C)** Northern blot analysis of the expression of *yggL* over time in wild-type (wt) *E. coli* MG1655 reveals constant expression while the cells are still growing ($OD_{600\,nm}$ of 0.4–2.0), whereas its expression is not detectable in late stationary phase ($OD_{600\,nm}$ of 2.0 + 4 h). All three *yggL*-containing transcript species (originating from TSS2, TSS3 and the processed transcript) can be detected. 5S rRNA served as loading control. **(D)**

of 0.4 is stronger than from TSS3, meaning that, during exponential growth, the bicistronic *trmI-yggL* mRNA is more expressed than the monocistronic *yggL* mRNA.

To experimentally validate the transcriptional organization of *yggL* and how it is expressed over time, a growing *E. coli* culture was sampled several times after dilution from the overnight culture, the RNA isolated and northern blotting performed (Figure 2.25 C). As expected from the dRNA-seq data, three distinct bands were detected with lengths of $\sim$1,200, $\sim$750 and $\sim$550 nt, which fit to the transcripts originating from TSS2, TSS3 and the cleavage site, each respectively. As observed in the dRNA-seq data, the longest band (*trmI-yggL*) disappeared after the culture reached an $OD_{600\,nm}$ of 2.0, whereas the two shorter bands (*yggL* and the processed transcript including *yggL*) did not change. During late stationary phase ($OD_{600\,nm}$ of 2.0 + 4 h), *yggL* could not be detected. Taken together, YggL was identified as a potential binder of the 50S subunit that is expressed in a σ^{70}-dependent manner.

2.9.3 YggL is highly conserved among γ-proteobacteria

YggL is the only member of a family of proteins containing a DUF469 (DUF = domain of unknown function), which, according to InterPro, is predicted to function as phosphotransferase (El-Gebali et al., 2019; Mitchell et al., 2019). To analyze how conserved YggL is, eggNOG (Huerta-Cepas et al., 2016) was searched for homologs, revealing 150 species containing *yggL*, the vast majority of which were represented by the class of γ-proteobacteria (Figure 2.26 A). A smaller number of homologs could further be identified in the closely related class of β-proteobacteria, among which presence of *yggL* is mostly limited to the orders of Burkholderiales and Neisseriales. Interestingly, amino acid conservation of YggL among γ-proteobacteria is very high (Figure 2.26 B), suggesting that its function in these bacteria is conserved as well.

2.9.4 Ribosome-binding properties of YggL

The Grad-seq protocol employed here uses a low magnesium concentration of 1 mM for both the lysate buffer and the glycerol gradient, which is known to partially dissociate 70S ribosomes into their subunits (Ron et al., 1968). YggL was observed both at the 50S subunit and in the pellet fraction (Figure 2.24), meaning there are several options that could have led to this observation: 1) YggL binds to the 50S subunit

A

B

Figure 2.26: Conservation of YggL. (**A**) Phylogenetic analysis of YggL based on 150 protein sequences deposited in eggNOG 4.5.1 (Huerta-Cepas et al., 2016) (COG3171). YggL homologs were only found in γ-proteobacteria (orange) and β-proteobacteria (blue). (**B**) Multiple sequence alignment of YggL homologs from *E. coli* (ECO), *Salmonella* Typhimurium (STM), *Salmonella bongori* (SBG), *Shigella flexneri* (SFL), *Klebsiella pneumoniae* (KPN), *Enterobacter cloacae* (ENT), *Citrobacter rodentium* (CRO), *Pantoea* spp. (PAN) and *Yersinia pestis* (YPE) using Clustal Omega (Madeira et al., 2019). YggL is highly conserved on amino acid level between different members of γ-proteobacteria. Residues with ≥50 % identity are highlighted in a blue gradient. Visualization was performed using Jalview (Waterhouse et al., 2009).

Figure 2.27: Verification of the YggL-50S subunit interaction. (**A**) Sucrose polysome gradient analysis using a *yggL*-3xFLAG strain shows sedimentation of YggL-3xFLAG together with the 70S ribosome. L, lysate (input control). (**B**) *In vitro* reconstitution of 100 (black) or 400 pmol (orange) purified YggL and 400 pmol of purified ribosomes obtained from a Δ*yggL* strain followed by sucrose polysome gradient analysis reveals binding of YggL to the 70S ribosomes.

by itself and to 50S subunits of assembled 70S ribosomes/polysomes. 2) YggL binds only to the 50S subunit of 70S ribosomes/polysomes and was subsequently detected at the 50S subunit due to partial dissociation of the whole ribosome. 3) YggL binds to the 50S subunits that are part of other, higher-order ribosomes such as 100S ribosomes (compare to Rmf discussed in 2.9.1) and was subsequently detected at the 50S subunit due to partial dissociation of the whole ribosome.

To test which of these possibilities is the underlying cause of the observed sedimentation of YggL, sucrose polysome gradients at 10 mM magnesium were performed with the *yggL*-3xFLAG strain. The higher magnesium concentration stabilizes 70S ribosomes and polysomes (Ron et al., 1968), which allows discrimination between free 50S subunits and ones assembled into 70S ribosomes. Western blotting of the resulting gradient fractions revealed YggL in the 70S monosome fraction (Figure 2.27 A). Importantly, no signal was detected in polysomal fractions and only minimal levels were detected in the 50S and bulk fractions. It is of note that the measured UV absorbance of the 50S subunit of the *yggL*-3xFLAG strain was higher than what is usually observed (compare to, *e.g.*, Figure 2.5 D).

To further examine whether YggL binds exclusively to 70S monosomes, *in vitro* reconstitution experiments using recombinantly purified YggL and purified ribosomes from a Δ*yggL* strain were performed. For this, 400 pmol of purified ribosomes were allowed to associate with 100 or 400 pmol of YggL and subsequently run on a sucrose gradient (Figure 2.27 B). When 100 pmol YggL were used, almost all (∼89 %, ∼89 pmol) of it could be detected at the 70S monosomes with only little signal at the top of the gradient, where free YggL was expected. No signal was observed at the 50S subunit or the polysomes. 400 pmol YggL showed the same distribution along the ribosomal peaks, whereas a lot of it was unbound at the top of the gradient. The bound fraction of YggL in this case was ∼18 % (∼72 pmol), which is very similar to the results obtained with 100 pmol YggL. These results suggest that YggL only binds to 50S subunits involved in 70S monosomes and that possibly only a subpopulation of 70S monosomes allows this association.

A

B
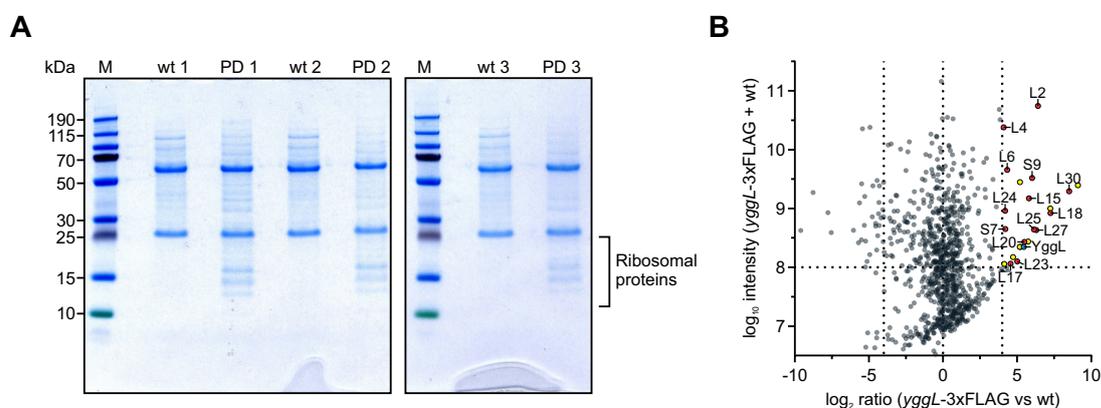
Figure 2.28: AP/MS of YggL-3xFLAG. (**A**) Coomassie-stained gels after affinity purification (AP) of *E. coli* wild type (wt) or *yggL*-3xFLAG (PD). Several proteins <25 kDa are enriched in the PD compared to the wt. M, size marker. (**B**) MS analysis of the proteins enriched in the PD *vs.* the wt. Apart from the expected enrichment of YggL, several proteins of the large ribosomal subunit as well as two of the small ribosomal subunit are enriched.

2.9.5 AP/MS validation of the ribosome-binding properties of YggL

As additional validation of the ribosome-binding properties of YggL, AP/MS (see 1.2.1) assays comparing the wild-type strain against the *yggL*-3xFLAG strain were performed. Compared to the wild type, a clear enrichment of small proteins <25 kDa was observed in the *yggL*-3xFLAG strain, which fit the size for ribosomal proteins (Figure 2.28 A). Identification of these proteins was performed by MS analysis, revealing the pull-down of many proteins of the large ribosomal subunit as well as S7 and S9 of the small ribosomal subunit (Figure 2.28 B). The pull-down of mostly 50S subunit proteins adds to the evidence that YggL specifically interacts with the large ribosomal subunit.

2.9.6 Deletion of *yggL* causes an increase in free 50S subunits

Based on its association with the ribosome, it seems obvious that YggL might somehow be involved in the process of translation. To test the consequences of deleting *yggL*, growth curve experiments were performed (Figure 2.29 A). The Δ*yggL* mutant did not show any differences in growth in rich media, suggesting that the function of YggL might either be rather limited or very specialized. Since the *yggL*-3xFLAG strain showed an increase in free 50S subunits in a polysome gradient (Figure 2.27 A), the wild-type, Δ*yggL* and Δ*yggL* complementation strains were run on a polysome
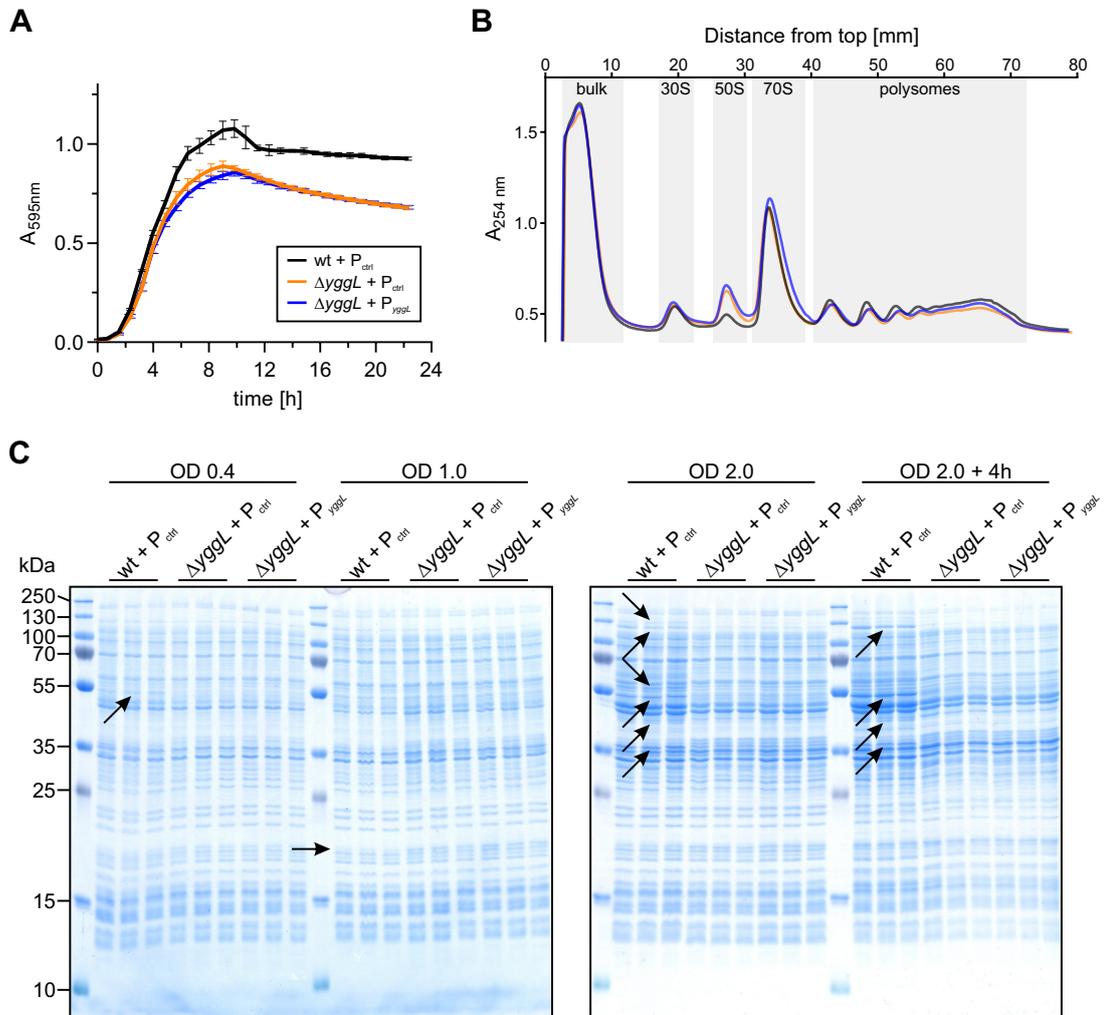
Figure 2.29: Phenotypes of the *yggL* deletion. (**A**) Growth curves of *E. coli* MG1655 wild type and Δ*yggL* show no growth defect of the knockout mutant. Error bars show SD from the mean, n = 3. (**B**) Sucrose polysome gradient analysis of *E. coli* MG1655 wild type, Δ*yggL* and a complementation strain. The knockout shows an increased amount of free 50S subunits, which is not complemented by constitutive overexpression of *yggL*. (**C**) Analysis of the total protein content of *E. coli* MG1655 wild type, Δ*yggL* and a complementation strain by SDS-PAGE. Several proteins are up- or downregulated in the knockout (highlighted by arrows), which is not complemented by constitutive overexpression of *yggL*.

gradient. As observed before for the *yggL*-3xFLAG strain, the knockout of *yggL* caused an increase in free 50S subunits and even a slight decrease in polysomes (Figure 2.29 B). This phenotype could, however, not be complemented by constitutive overexpression of *yggL*.

To test whether the observed differences in ribosome abundance have an influence on translation of specific proteins, total protein was collected in a time course experiment and separated via PAGE (Figure 2.29 C). Interestingly, several bands increased or decreased when comparing the wild type to the knockout strain. The strongest change at early exponential phase ($OD_{600\,nm}$ of 0.4) was a single band at around 55 kDa that disappeared in the knockout. At mid-exponential phase ($OD_{600\,nm}$ of 1.0), a $\sim$20 kDa band was absent from the wild type but present in the knockout. In contrast, six proteins at $\sim$130, $\sim$125, $\sim$58, $\sim$54, $\sim$45 and $\sim$36 kDa were more expressed in the wild type than in the knockout at early stationary phase ($OD_{600\,nm}$ of 2.0). Surprisingly, except for the bands at $\sim$130 and $\sim$58 kDa, the same was observed at late stationary phase ($OD_{600\,nm}$ of 2.0 + 4 h), even though *yggL* is not expressed anymore at this point of growth (Figure 2.25 C).

Additionally, in all tested conditions, several bands slightly heavier than 35 kDa were upregulated in the knockout. As observed in the polysome gradient, overexpression of *yggL* could not rescue the phenotype in any of the tested conditions. Attempts to express *yggL* from a low copy plasmid under its native promoter did not show any mRNA expression, possibly due to inhibition by chloramphenicol, since the same behavior was also observed for the control plasmid (data not shown). Together, these results show that deletion of *yggL* increases the number of free 50S subunits and changes expression of several proteins.

2.10 Discussion

Being the most popular bacterial model organism, a wide array of studies has investigated the interactome of *E. coli* using both classical low-throughput and global high-throughput methods. However, these studies mostly focused on the interaction between proteins, excluding RNPs from the global approaches. With the introduction of Grad-seq (Smirnov et al., 2016), it was for the first time possible to obtain

information about the *in vivo* complexome of both RNA and protein from a single experiment.

In this chapter, Grad-seq was employed in order to draft a global map of RNA and protein interactions in *E. coli*. After several rounds of optimization (see 2.1), highly reproducible gradients were obtained (Figures 2.4 and 2.5), whose coarse-grained analysis using standard biochemical assays already revealed several surprises: Most of the tested sRNAs showed a wide spread throughout the gradient with peaks around fraction 5 and in the pellet, suggesting Hfq and ribosome association, respectively (Figure 2.5 A and B). Importantly, this behavior could be reproduced using sucrose polysome gradients (Figure 2.5 D) and was further revealed on a global scale in the RNA-seq data (Figure 2.11 B). Several scenarios can lead to ribosome association of sRNAs:

1. The sRNA is dual-function RNA, meaning that it has both regulatory functions as an sRNA and encodes a protein. Translation of the sRNA in this case would cause ribosome association. The best-studied example of a dual-function RNA is the quorum sensing-induced *Staphylococcus aureus* RNAIII, which negatively regulates translation of several major surface proteins and positively regulates translation of two more proteins, including α-hemolysin (Bronesky et al., 2016). At the same time, it encodes the *hld* gene, the CDS of δ-hemolysin (Bronesky et al., 2016). In *E. coli*, the dual-function RNA SgrS encodes *sgrT*, both of which act in concert to combat glucose-phosphate stress (Lloyd et al., 2017; Wadler and Vanderpool, 2007). SgrS was mostly detected in the pellet fraction of the gradient (Figure 2.15), suggesting that the conditions used here might favor translation of SgrT. SgrT could not be detected in the MS data, which is not surprising, however, given that it is only ~5.3 kDa and small proteins are notoriously difficult to detect in MS (in fact, the smallest protein detected here was RpmH, a ~5.4 kDa ribosomal protein for which it can safely be assumed that it is way more abundant than SgrT). Translation activation of *yigL* by SgrS (Papenfort et al., 2013; Sun and Vanderpool, 2013) might further cause ribosome association of SgrS (see next point).

2. Binding of the sRNA positively regulates translation of its target, thereby shifting it to the pellet. A classic example of this type of regulation is DsrA (Papenfort and Vanderpool, 2015), which functions via an anti-antisense mechanism, opening up an inhibitory secondary structure in the 5′ UTR of *rpoS* and therefore allowing ribosome binding and translation (Lease et al., 1998; Majdalani et al., 1998). RpoS is an alternative σ-factor that mediates the transition from logarithmic to stationary growth in *E. coli* (Battesti et al., 2011), meaning that in the conditions used here (early stationary phase), RpoS translation should be active and indeed, DsrA was mostly detected in the pellet (Figure 2.15).

3. Binding of the sRNA negatively regulates a cistron within a polycistronic mRNA, whose other genes still can be translated, causing the sRNA to co-sediment with the ribosomes. Generally, inhibition of translation of a cistron leads to the degradation of the whole mRNA (Balasubramanian and Vanderpool, 2013). This is, however, not always the case: Spot 42, for example, specifically downregulates translation of *galK* within the *galETKM* mRNA, which does not cause degradation of the mRNA (Møller et al., 2002). Grad-seq revealed Spot 42 to be mostly present in the pellet fraction (Figure 2.5 A), which to some extent can be explained by the described behavior. It has to be mentioned, though, that Spot 42 also possesses an ORF (Yano, Chao and Vogel, unpublished; Gisela Storz, personal communication), meaning that it can be translated, which at least partially should be responsible for its ribosome association too.

Other scenarios such as sRNA-dependent target activation via inhibition of mRNA degradation, *e.g.*, translational activation of *cfa* by RydC (Fröhlich et al., 2013), or direct interaction between ncRNAs and the ribosome as observed in eukaryotes (Pircher et al., 2014) could also shift an sRNA to the pellet. Crucially, Hfq is generally needed for the mentioned regulatory mechanisms. In agreement with this, Hfq exhibited high abundance in the pellet of the gradient too (Figure 2.15). In *Salmonella* Grad-seq, both sRNAs and Hfq showed a peak around fraction 5 and an additional increase in abundance in fraction 20 (Smirnov et al., 2016), showing the

same trend as observed for *E. coli*. However, the pellet fraction was not analyzed in the *Salmonella* study and the ribosome association of sRNAs and Hfq can therefore not be safely estimated.

Similar to Hfq, ProQ also was abundant in the ribosome-associated pellet fraction (Figure 2.15). Importantly, in *Salmonella*, ProQ showed a strong increase in abundance toward the bottom of the gradient as well (Smirnov et al., 2016). Yet, in contrast to Hfq-associated sRNAs, the ProQ-associated sRNAs sedimented mostly in early fractions and less so in the pellet (Figure 2.11 F). Since the exact functions of ProQ and its associated sRNAs are less understood than those of Hfq, it can only be speculated what the reason for this observation might be. On the one hand, ProQ is known to bind and stabilize a large array of often highly structured sRNAs (Holmqvist et al., 2018; Melamed et al., 2020; Smirnov et al., 2016), explaining its sedimentation around fraction 4, where many of its associated sRNAs peaked (Figure 2.11 F) and which was previously observed in *Salmonella* (Smirnov et al., 2016). On the other hand, ProQ possibly facilitates base pairing between sRNAs and their targets, as demonstrated for *Salmonella* RaiZ inhibiting translation of *hupA* (Smirnov et al., 2017b).

As discussed for Hfq-dependent sRNAs above, formation of such ternary inhibitory complexes within polycistronic mRNAs could cause ProQ to sediment together with ribosomes. Although not shown yet, it is possible that ProQ can perhaps also form translation-permitting complexes with sRNAs and their targets, which again would cause ribosome association. Finally, ProQ was shown to bind the 3′ UTRs of a vast number of mRNAs and thereby protect them from degradation by RNase II in *Salmonella* (Holmqvist et al., 2018) or PNPase in *Neisseria* (Bauriedl et al., 2020). Provided continued association to its mRNA targets while upstream translation is ongoing, this ProQ-dependent protection currently seems the most plausible option for the observed ribosome association of ProQ. Based on the sedimentation behaviors of sRNAs and their associated proteins defined in this study, it is now possible to estimate their preferred macromolecular state, *i.e.*, their most stable complex.

In addition to globally displaying the sedimentation behavior of RNAs of known function, the Grad-seq data obtained here led to the discovery that the functionally

unknown sRNA RyeG in fact is a prophage-derived mRNA encoding a toxic small protein (see 2.8). Most sRNAs showed sedimentation in the early fractions and the pellet (Figure 2.11 B). RyeG was strikingly different: it co-sedimented with the 30S subunit and showed some additional abundance in the pellet (Figure 2.17).

The existence of RyeG was first predicted in a biocomputational study and named "IS118" (Chen et al., 2002). It is encoded in the cryptic (defective) prophage CPS-53 (a.k.a. KplE1), which is one of 9 prophage regions of *E. coli* K-12 (Wang et al., 2010). CPS-53 is defective for lysis but can still perform excisive recombination (Champ et al., 2011; Panis et al., 2007, 2010b,a, 2012; Puvirajesinghe et al., 2012). Not much is known about the function of its genes and why it has been kept in the chromosome. Often, domestication of prophages confers positive traits to their hosts, for instance, stress resistance or resistance against other phages (Bobay et al., 2014). This is also the case for CPS-53, for which one study showed involvement in resistance against acid and oxidative stress (Wang et al., 2010). Another study overexpressed the *yfdQRST* gene cluster of CPS-53, leading to inhibition of the initiation of chromosomal replication, which was suggested to be beneficial under certain stress conditions (Noguchi and Katayama, 2016). Only one study investigated potential effects of RyeG so far: its overexpression led to reduction in biofilm production, swarming and swimming motility as well as type I and curli fimbriae production (Bak et al., 2015). The authors showed no detailed data on growth defects related to overexpression of RyeG but it is noteworthy that from the 99 sRNAs they tested, RyeG was the only one to negatively affect all of the tested processes (Bak et al., 2015).

In the present study, overexpression of RyeG showed toxic effects on cell growth in both *E. coli* K-12 as well as other *E. coli* and *Salmonella* strains that do not carry the CPS-53 prophage (Figure 2.20), implying that the observed phenotype is independent of other genes of CPS-53. The toxicity manifested itself through drastically prolonged lag times in all investigated strains. Lag phase is defined as the time it takes a bacterium until it first divides after being refreshed in new medium, meaning that it is the phase that prepares for growth (Bertrand, 2019). Increased lag times can be detrimental to a bacterial culture because it might be less fit and be outcompeted by other, faster growing bacteria.

However, increased lag times can also be beneficial: in *S. pneumoniae*, serotypes with a greater potential for invasive disease were shown to have longer lag phases, possibly due to increased immune evasion (Bättig et al., 2006). Further, increased lag times can confer resistance against antibiotics (Bertrand, 2019), as impressively demonstrated in a study in *E. coli* showing that ampicillin-tolerant bacteria were also tolerant to norfloxacin (whose mode of action is different than that of ampicillin) due to prolonged lag times (Fridman et al., 2014). The authors termed this phenomenon "tolerance by lag" (Fridman et al., 2014). In the case of overexpression of RyeG, the bacteria appeared to grow rather normally after overcoming the lag phase (Figure 2.20 A–D). Together with the increased stress resistance provided by CPS-53 (discussed above), this toxic effect, *i.e.*, the increased lag phase, of RyeG might also be beneficial under certain circumstances. However, no conditions were identified under which such a positive effect could be observed.

As already mentioned, RyeG was of outstanding interest because of its putative 30S subunit association. The only sRNA previously shown to directly interact with the 30S subunit was the *rpoS*-activating DsrA (discussed above) (Worhunsky et al., 2003). However, the interaction of the 30S subunit with RNA generally is thought to be unique to mRNAs, which interact with 30S subunits in order to form translation initiation complexes, the first and rate-limiting step in protein biosynthesis (Gualerzi and Pon, 2015). Since RyeG is specific to *E. coli* K-12 (discussed above), no predictions on putative conserved ORFs could be performed and therefore a naïve approach considering all potential ORFs within RyeG was taken in order to test whether RyeG is indeed noncoding or whether it actually encodes a protein. 30S toeprinting assays and mutational studies demonstrated that RyeG includes an ORF here termed ORF2 that encodes a 48 aa long small protein that apparently is solely responsible for the observed phenotype of RyeG overexpression (Figure 2.22 B).

While this study was still ongoing, others identified ORF2 using a modified ribosome profiling approach and renamed it *yodE* without investigating its functions (Weaver et al., 2019). Homology searches for the protein sequence of ORF2/*yodE* did not reveal any known similar sequences, which therefore did not allow functional predictions. The mechanism of action of prolonging the lag phase by RyeG/YodE will hence need further studies.

Apart from investigating interactions of RNAs and their cognate RBPs, Grad-seq also enables the global analysis of stable *in vivo* PPIs. As discussed in 1.2, many of the previous studies on the global interactome of *E. coli* relied on massive epitope-tagging of single proteins in order to generate binary interactome maps (Arifuzzaman et al., 2006; Babu et al., 2018; Butland et al., 2005; Hu et al., 2009; Rajagopala et al., 2014). In contrast, a wild-type strain was used to generate the Grad-seq maps in this thesis, which revealed the protein interactome based on a single experiment. Quality control analyses revealed that the subunits of the majority of heterocomplexes, for which all subunits were detected in the MS dataset, showed high correlation, indicating intact complexes (Figure 2.14 A). This could further be observed for conserved protein-protein complexes ranging from small (succinyl-CoA synthetase, $\sim$140 kDa) to large (50S subunit, $\sim$1.6 MDa) complexes (Figures 2.13 E and 2.14 B), demonstrating the large gamut of resolved complexes in this Grad-seq resource.

However, the predicted cellular localization of the detected proteins also indicated one of the limitations of Grad-seq: Cytosolic proteins were enriched, disfavoring membrane-associated proteins (Figure 2.13 A and B, discussed in 1.2.5.9). In comparison, global AP/MS does not rely on the solubility of the bait proteins and was recently successfully applied to analyze membrane complexes (Babu et al., 2018). One option to enable the investigation of such complexes using Grad-seq is the utilization of detergents that keep membrane complexes in solution, as was recently shown in a study employing gradient-based PCP (see 1.2.4.3; (Carlson et al., 2019)).

One of the main ways of using Grad-seq as a resource is to look up the sedimentation profile of a protein (or RNA) of interest in order to obtain a quick overview whether this particular protein might be involved in a complex. The assumption here is that one can estimate the sedimentation coefficient of the protein of interest (see 1.2.5.4; (Erickson, 2009)) and therefore predict in which region of the gradient it should sediment if it would not be involved in a complex (this usually is the very top of the gradient). To showcase this unique feature of Grad-seq, a search for small proteins that are highly likely to form complexes was conducted (Figure 2.23). Of

the 97 identified small proteins, the functionally uncharacterized YggL was chosen for further evaluation.

YggL sedimented mostly at the 50S subunit and in the pellet fraction, suggesting ribosome association (Figure 2.24). Intriguingly, the same behavior was previously observed for *Salmonella* YggL (Smirnov et al., 2016), which hinted at a conserved function of the protein. Follow-up experiments then showed that the target of YggL in fact is the 70S ribosome (Figure 2.27), suggesting that the observed 50S subunit association possibly could be attributed to partial disassembly of 70S ribosomes caused by the low magnesium concentration applied here (Ron et al., 1968). Interestingly, AP/MS (see 1.2.1) of a *yggL*-3xFLAG strain resulted in pull-down of 12 strongly enriched proteins of the large ribosomal subunit, namely L2, L4, L6, L15, L17, L18, L20, L23, L24, L25, L27 and L30 (Figure 2.28 B). A striking commonality between these proteins is that, except for L2, all of them are on the cytosolic side of the 50S subunit, facing away from the inter-subunit interface between the 30S and 50S subunits (Nikolay et al., 2015). While it seems unlikely that the rather small (108 aa) YggL can interact with all of them, this observation provides some evidence that YggL might bind the cytosol-oriented side of the 50S subunit. Further, L20, L24, L27 and L30 are not found within 50S subunit precursor particles in *E. coli* (Pokkunuri and Champney, 2007), which is in agreement with the model that YggL binds to assembled 70S ribosomes.

Knockout of *yggL* did not cause any growth defect but instead led to an increase in free 50S subunits (Figure 2.29 A and B). A previous study implicated YggL in 70S ribosome assembly (Chen and Williamson, 2013), suggesting that the increase in free 50S subunits might be due to decreased assembly of 70S ribosomes. If true, this should also result in higher amounts of free 30S subunits, which could not be observed. At the same time, lower amounts of 70S ribosomes and/or polysomes would be expected. Indeed, a slight decrease in polysome levels was measured in the *yggL* knockout but it remains unclear whether this difference is meaningful. Importantly, the 50S subunit peaks fully overlapped in the wild type and the knockout strain, indicating that fully assembled subunits accumulated since precursor particles such as the 45S precursor exhibit slower sedimentation in the gradient (Jomaa et al., 2014). Attempts to complement this phenotype from a plasmid failed, which could be due

to polar effects of the knockout strain. However, to exclude polar effects, the *yggL* knockout was carefully designed to only delete the N-terminal half of the gene, keeping all TSSs of this complex locus (Figure 2.25 A) intact. It is therefore possible that the stoichiometry of YggL within the cell is an important factor to ensure its proper function (more on copy numbers below).

Transcription of *yggL* occurs in a σ^{70}-dependent manner (Figure 2.25 B and C), which might point toward a function in the assembly of ribosomes as discussed above or a function that is related to translation of specific proteins. In the latter case, proteins such as elongation factor P, which is essential to prevent translational stalling caused by polyproline stretches (Doerfel et al., 2013; Ude et al., 2013), are involved in aiding translation of certain proteins or amino acid sequences that could otherwise not be translated optimally. Of such proteins, the chaperone trigger factor is the most-studied one that binds to the cytosol-facing side of the 50S subunit—the predicted binding site of YggL—and which is involved in binding and folding of nascent polypeptide chains exiting the ribosome (Bhandari and Houry, 2015; Hoffmann et al., 2010).

YggL expression ranges from ~1,300–6,700 copies per *E. coli* K-12 cell (Li et al., 2014; Schmidt et al., 2016), which face ~8,000–70,000 ribosomes per growing cell (Bremer and Dennis, 2008). Based on this ~10-fold excess of ribosomes over YggL, a general function like the one of trigger factor therefore seems unlikely and might explain the narrow spectrum of bacterial species harboring the *yggL* gene (Figure 2.26 A). Still, inspection of the total protein content of a *yggL* knockout strain highlighted several changes in the abundance of proteins when compared to the wild type (2.29 C).

Homology searches of YggL did not result in any hits with known function. In fact, YggL forms its own protein family possessing the DUF469 (El-Gebali et al., 2019), thereby preventing functional predictions. In the future, cryogenic electron microscopy of YggL-70S ribosome complexes will be helpful to pinpoint the binding site of YggL and with that build the foundation of understanding its exact molecular function.

In summary, the results of this chapter provide a valuable RNA/protein complexome resource for *E. coli* MG1655 that adds confidence to the wealth of previous

binary interactome studies (Arifuzzaman et al., 2006; Babu et al., 2018; Butland et al., 2005; Hu et al., 2009; Rajagopala et al., 2014). The dataset should lend itself to the analysis of related bacteria as well, many of which are important human pathogens (Kersters et al., 2006; Rizzatti et al., 2017). Furthermore, the optimized Grad-seq protocol presented here should generally be applicable to other bacteria and even eukaryotes, making Grad-seq an excellent option to obtain a quick overview of the complexome. Finally, based on the unique features offered by Grad-seq, two ribosome-associated molecules, RyeG and YggL, were detected and established as a prophage-derived toxic mRNA and a 50S subunit-binding protein, respectively.

Chapter 3

Grad-seq of *Streptococcus pneumoniae*

The majority of the work presented in this chapter has previously been published elsewhere:

- Hör, J., Garriss, G., Di Giorgio, S., Hack, L-M., Vanselow, J. T., Förstner, K. U., Schlosser, A., Henriques-Normark, B., and Vogel, J. (2020). Grad-seq in a Gram-positive bacterium reveals exonucleolytic sRNA activation in competence control. *EMBO J*, doi:10.15252/embj.2019103852

- Hack, L-M. (2019). Characterization of Cbf1 in *S. pneumoniae*. Master's thesis, University of Würzburg, Würzburg, Germany.

Further, parts of the work in this chapter have been performed in collaboration with the following people:

- Dr. Geneviève Garriss (Karolinska Institutet, Stockholm, Sweden) constructed all strains and performed the competence assays and growth curve experiments.

- Silvia Di Giorgio (ZB MED, Cologne), Dr. Jens Vanselow (Rudolf Virchow Center, Würzburg) and Dr. Thorsten Bischler (Core Unit SysMed, Würzburg) performed parts of the high-throughput data analyses.

- Lisa Marie-Hack performed parts of the experiments surrounding the characterization of Cbf1 during her Master's thesis and subsequent employment as technical assistant under supervision of Jens Hör and Prof. Dr. Jörg Vogel.

3.1   Grad-seq can be applied to a Gram-positive organism

Gram-positive bacteria, many of which are important pathogens (Woodford and Livermore, 2009), make ample use of post-transcriptional gene regulation by sRNAs (Brantl and Brückner, 2014; Desgranges et al., 2019; Quereda and Cossart, 2017; Wagner and Romby, 2015; Wassarman, 2018). Yet, in comparison to Gram-negatives, the molecular mechanisms of these processes remain largely unknown. This is especially true for organisms like the clinically highly relevant pneumococcus, which does not possess homologs of the well-known global regulators CsrA, Hfq or ProQ (Tettelin et al., 2001). Therefore, the pneumococcus was chosen as a model Gram-positive bacterium for the application of Grad-seq in order to learn more about its complexome and potentially identify novel RBPs. The results of this study should further be applicable to related organisms.

To test whether Grad-seq can be used for the pneumococcus, the optimized Grad-seq protocol (see 2.1) was applied to a culture of strain TIGR4 grown to mid-exponential phase in rich medium ($OD_{600\,nm}$ of 0.5). The only difference to the *E. coli* protocol was the lysis conditions, which had to be changed in order to allow proper lysis of the Gram-positive pneumococcus. The UV profile of the gradient (Figure 3.1 A) closely resembled the *E. coli* profile (Figure 2.4), revealing an early bulk peak and two peaks for the 30S and 50S subunits. Interestingly, RNA extraction from fractions around the 50S subunit did not cause any distortions, leading to an RNA concentration profile that was fully congruent with the UV profile (Figure 3.1 A).

Next, RNA gel analysis was performed to test whether the observed peaks fit with their corresponding house-keeping ncRNAs (Figure 3.1 B). As expected, tRNAs and the 16S and 5S/23S rRNAs were detected around fractions 3, 11 and 15, respectively, confirming that the peaks observed in the UV profile derived from these abundant ncRNAs. Several additional abundant bands were visible in the RNA gel, which could be attributed to M1 RNA (the RNase P RNA (Mondragón, 2013)), tmRNA
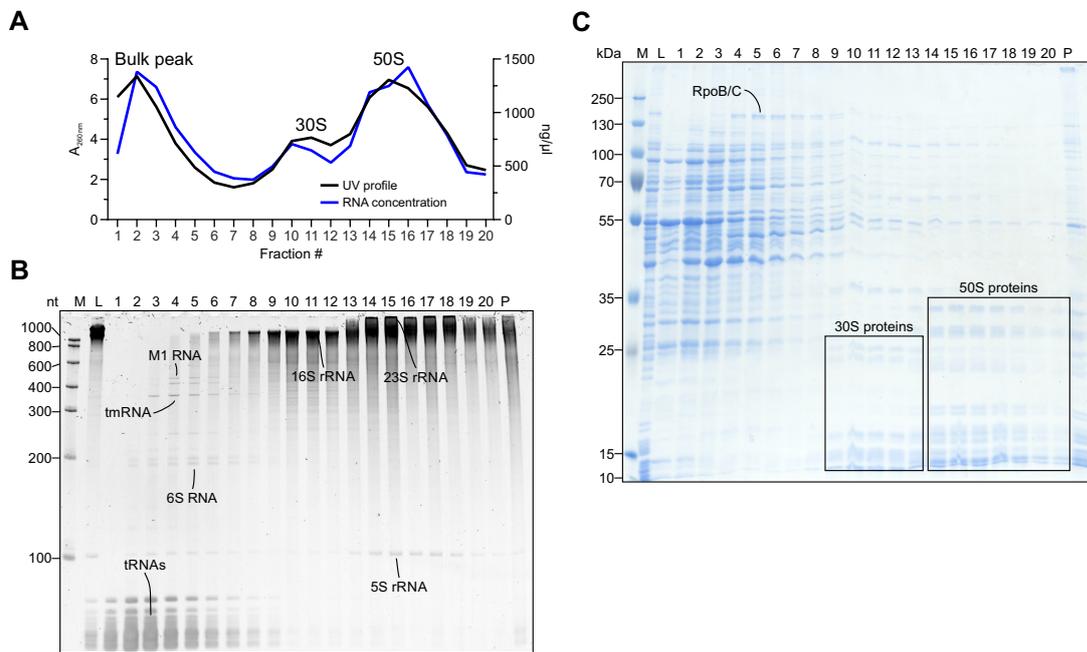
Figure 3.1: Overview of *S. pneumoniae* Grad-seq. (**A**) The general gradient profile monitored by UV measurement and the RNA concentration after RNA purification show congruent profiles and reveal a bulk peak and two peaks for the 30S subunit (around fraction 10) and the 50S subunit (around fraction 15). (**B**) RNA gel analysis reveals that the obtained RNA is of high quality and that several classes of abundant ncRNAs are readily visible. (**C**) SDS-PAGE of the protein content of the gradient reveals specific sedimentation patterns for all proteins and shows that the ribosomal proteins migrate in the same fractions as their corresponding rRNAs (shown in (**B**)), indicating intact complexes.

A

B
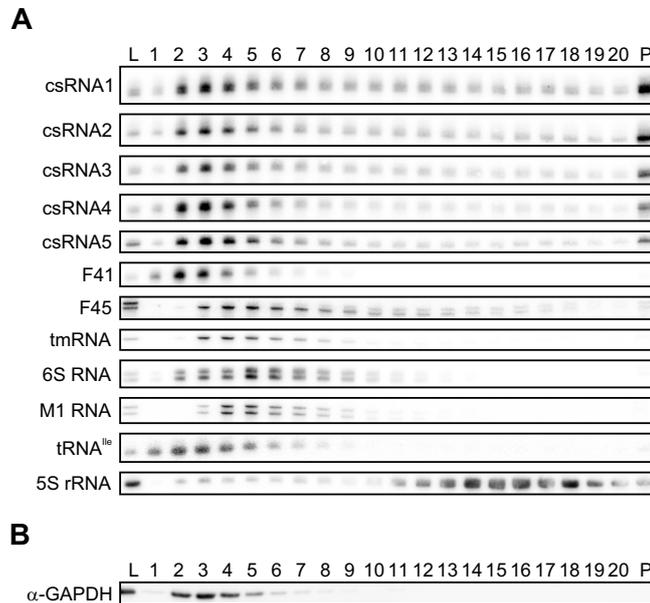
Figure 3.2: Overview of the sedimentation of specific molecules. (**A**) Northern blot analysis of different pneumococcal ncRNAs reveals different sedimentation profiles for many of them. Interestingly, the competence-inhibiting csRNAs all co-migrate and peak around fraction 3. (**B**) Western blotting of GAPDH shows a peak in fraction 3 and no signal toward the bottom of the gradient.

and 6S RNA. Similar to the RNA gel, the protein content of the gradient was analyzed by SDS-PAGE to get a rough overview of the protein sedimentation (Figure 3.1 C). RNAP sedimented as a particle of similar size as observed for *E. coli* (Figure 2.4), fitting with the profile of 6S RNA as its interacting RNA (Figure 3.1 B and C) (Wassarman, 2018). Finally, the ribosomal proteins could readily be detected in the fractions expected according to the UV and RNA analysis (Figure 3.1 A–C).

To gain a more detailed insight into the sedimentation behavior of ncRNAs, complementary northern blots were performed, revealing vastly different profiles for different classes of ncRNAs (Figure 3.2 A). Importantly, the previously observed patterns of tmRNA, 6S RNA, M1 RNA, tRNAs and 5S rRNA could be verified. The functionally redundant cia-dependent sRNAs (csRNAs) (Laux et al., 2015) are highly similar in sequence and structure but span a length from 93 nt (csRNA1) to 148 nt (csRNA5), meaning that their molecular weights drastically differ. Intriguingly, they still co-migrated within the gradient and peaked in fraction 3 as well as the pellet, indicating they could be associated to the same RBP (Smirnov et al., 2017a). The functionally unknown F41 had its peak abundance around fraction 3 as well,

A

B

C
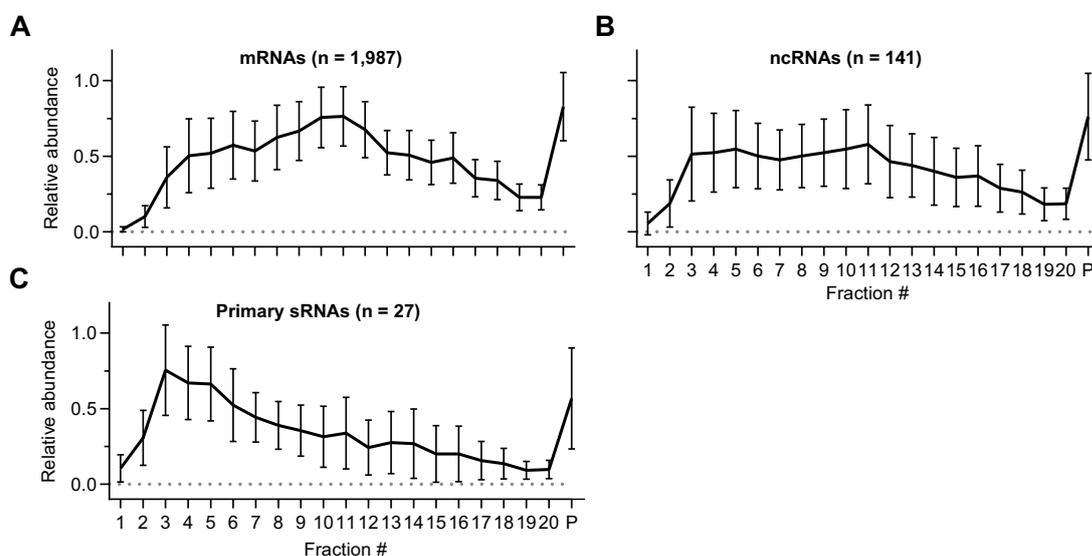
Figure 3.3: Sedimentation profiles of different pneumococcal RNA classes. The average sedimentation profiles of mRNAs (**A**), ncRNAs (**B**) and primary sRNAs (**C**) reveal that different RNA classes show different gradient sedimentation behavior. Profiles are normalized to a maximum of 1. Error bars show SD from the mean.

whereas F45, whose function is also unknown, formed a particle of similar size as the tmRNA RNP. Finally, western blotting revealed glyceraldehyde-3-phosphate dehydrogenase (GAPDH) to peak around fraction 3 (Figure 3.2 B). Taken together, these results demonstrate that Grad-seq is able to resolve complexes from the Gram-positive pneumococcus without major changes to the protocol.

3.2 Overview of the pneumococcus Grad-seq data

3.2.1 Overview of the RNA-seq data

RNA-seq of the 20 fractions plus the pellet resulted in 2,240 transcripts passing the filter of ≥ 100 reads over all fractions. This accounted for $\sim 88\%$ of the annotated transcriptome and included 1,987 mRNAs, 141 ncRNAs, 42 riboswitches as well as all tRNAs and rRNAs. Similar to what was observed for *E. coli* (Figure 2.11 D), mRNAs populated the whole gradient and peaked in the pellet, where 70S ribosomes were found (Figure 3.3 A). ncRNAs on the other hand sedimented more broadly than the mRNAs (Figure 3.3 B), which is in contrast to the *E. coli* ncRNAs, which mostly peaked in early fractions (Figure 2.11 B). The annotation of *S. pneu-*

A

Localization of the total *S. pneumoniae* proteome

Total = 2,113

B

Localization of the *S. pneumoniae* Grad-seq proteome

Total = 1,309
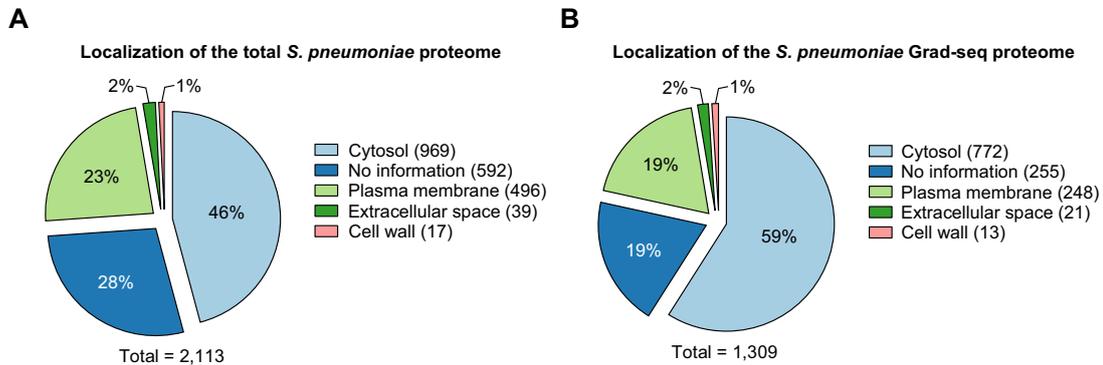
Figure 3.4: Grad-seq enriches cytosolic proteins. (A, B) Comparison of the total and the Grad-seq proteomes reveals enrichment of cytosolic proteins in the Grad-seq experiment. Note that the total number of proteins shown here is different from the number of detected proteins because some proteins have more than one localization assigned. Localization prediction is based on BioCyc (Karp et al., 2019).

moniae used here (Warrier et al., 2018) contains ncRNAs from several independent studies and many of them were only predicted and not validated (Acebo et al., 2012; Kumar et al., 2010; Mann et al., 2012), meaning that some of them might not be true ncRNA candidates.

To test whether a subset of true sRNAs shows a different average profile than what was observed for all ncRNAs (Figure 3.3 B), a dRNA-seq dataset that allows the identification of primary transcripts (collaboration between the Vogel and Henriques-Normark laboratories, unpublished) was used to identify independently transcribed sRNAs. All of the 27 identified primary sRNAs contained strong TSSs and ρ-independent terminators and were therefore considered true sRNAs. Interestingly, these 27 sRNAs showed an average sedimentation profile with a peak around fraction 3 that suggested their involvement in small complexes (Figure 3.3 C) and which was reminiscent of the data obtained for the *E. coli* ncRNAs (Figure 2.11 B). Conversely, this implies that many of the 141 ncRNAs found in the pneumococcus dataset might not represent true sRNAs or that they are involved in RNPs of greatly differing sizes, causing them to spread throughout the gradient.

3.2.2 Overview of the combined RNA-seq and MS data

Parallel MS analysis of each fraction detected 1,301 proteins within the gradient, representing ∼62 % of the proteome as annotated in UniProt (The UniProt Consortium,

2019). As observed for *E. coli* (Figure 2.13 B), the Grad-seq protocol enriched soluble pneumococcal proteins in comparison to the total proteome (Figure 3.4 A and B). This enrichment of cytosolic proteins could be attributed to a depletion of membrane-associated proteins as well as proteins with no information about their localization based on BioCyc predictions (Karp et al., 2019). Consequently, pneumococcal proteins with unknown cellular localization are likely membrane-associated, form aggregates or are lowly expressed in the conditions used here.

To find out more about the potential involvement of RBPs in stable cellular complexes, the MS dataset was cross-referenced with a list of 78 RBPs as predicted by UniProt (The UniProt Consortium, 2019). 77 of the 78 proteins were detected in the gradient and were assigned to functional classes of RBPs (Figure 3.5). As expected, proteins involved in tRNA modification such as aminoacyl-tRNA synthetases (*e.g.*, TyrS) or the CCA-adding enzyme (Cca) co-migrated with tRNAs. In contrast, RNases did not show a common profile in the gradient and RNA modification enzymes showed a trend of co-migration with ribosomal subunits. SP_0776 is the KhpA subunit of the KhpA/B heterodimeric RBP discovered to bind mRNAs, sRNAs and tRNAs in the pneumococcus (Winther et al., 2019; Zheng et al., 2017). While SP_0776 (KhpA) was among the predicted RBPs, SP_2040 (KhpB) was not in this list and therefore added to Figure 3.5 for clarity. They both exhibited congruent gradient profiles and peaked around fraction 3, indicating they indeed form a complex and that they might further form stable complexes with tRNAs and primary sRNAs, which were found to show similar sedimentation (Figures 3.1 B and 3.3 C).

While the knowledge on pneumococcal complexes is inferior to the knowledge on *E. coli* complexes, there still are many conserved multimeric enzymes present. The cytosolic F_1 complex of the ATP synthase, for example, is a five-subunit complex (Junge and Nelson, 2015) that was readily reproduced by Grad-seq (Figure 3.6 A). The subunits also showed elevated abundance in the pellet fraction, suggesting that it partially precipitated with the membrane-bound F_0 complex which it is in contact with (Junge and Nelson, 2015). Other examples of intact complexes included carbamoyl-phosphate synthase, glutamyl-tRNA^{Gln} amidotransferase (which co-migrated with tRNAs) and ribonucleoside-diphosphate reductase.
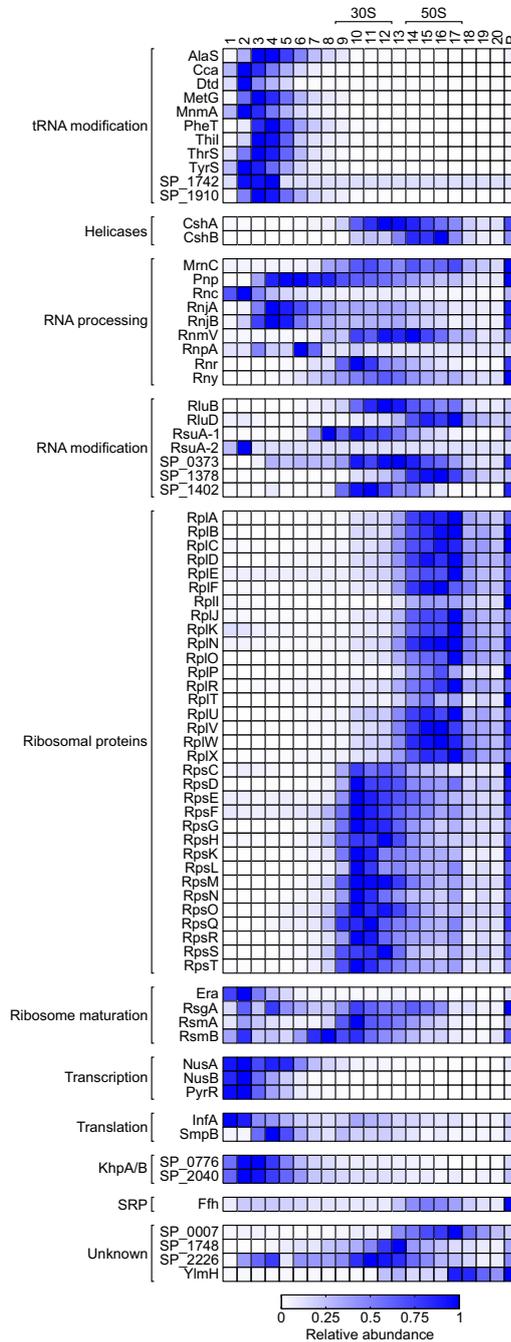
Figure 3.5: Overview of the sedimentation of RBPs. Heat map showing the sedimentation profiles of all detected RBPs based on classification by UniProt (The UniProt Consortium, 2019). SP_2040 was not in the list of RBPs in UniProt, but was added based on its interaction with SP_0776 (Zheng et al., 2017; Winther et al., 2019). SRP, signal recognition particle. Profiles are normalized to a maximum of 1.

A

B
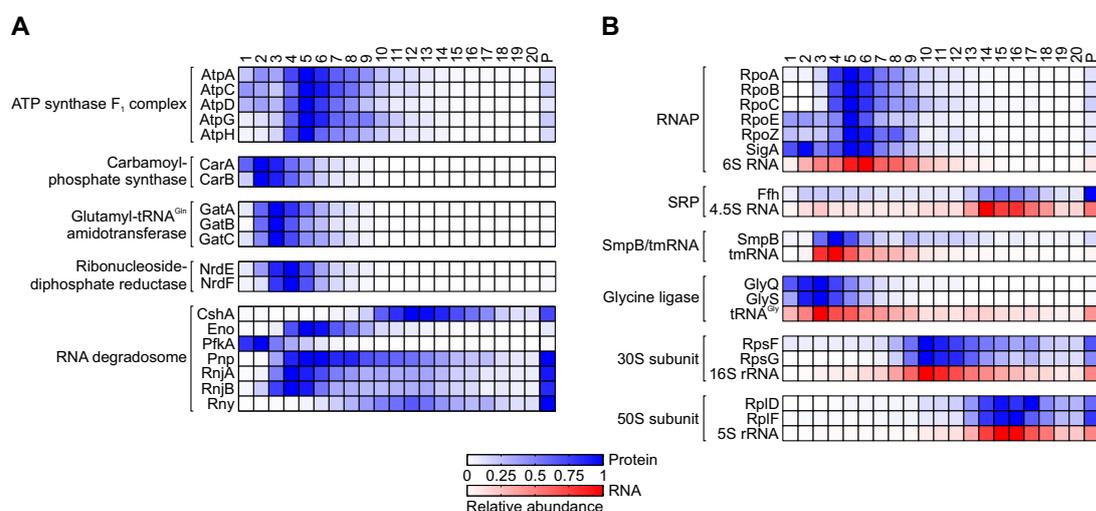
Figure 3.6: Overview of major protein complexes and RNPs. (**A**) Heat map of major intact protein complexes as well as the RNA degradosome. The subunits show excellent correlation, indicating intact complexes. (**B**) Heat map of major RNPs. The RNA and protein subunits of these RNPs show strong correlation, indicating that pneumococcal Grad-seq reveals intact RNPs. Profiles are normalized to a maximum of 1.

In comparison to Gram-negative organisms, Gram-positives contain a very different set of RNases (Durand and Condon, 2018). Similar to the enterobacterial RNA degradosome mentioned in 2.6.1, Gram-positives have been suggested to form RNA degradosomes too (Cho, 2017; Durand and Condon, 2018). In *Bacillus subtilis*, it consists of the RNA helicase CshA, enolase (Eno), phosphofructokinase (PfkA), the 3'→5' exonuclease PNPase (Pnp), the 5'→3' exonucleases RNase J1 (RnjA) and J2 (RnjB) as well as the endonuclease RNase Y (Rny) (Cho, 2017; Durand and Condon, 2018). The pneumococcus encodes for homologs of all of these proteins but their interactions have not been studied so far.

In the current Grad-seq dataset, all of the potential subunits were detected and, similar to what was observed for *E. coli* (Figure 2.13 E), no immediate co-sedimentation was found (Figure 3.6 A). However, all components except for PNPase and the two metabolic enzymes Eno and PfkA showed a peak around fractions 12–13. PNPase did not peak in these fractions but still exhibited high abundance (∼50 % of its peak intensity). Finally, for CshA and the four RNases, a strong intensity was detected in the pellet fraction. Especially RNase Y, which, similarly to RNase E in Gram-negatives (Mohanty and Kushner, 2018), is membrane-anchored and is as-

sumed to be the scaffolding protein in a potential degradosome complex (Cho, 2017; Durand and Condon, 2018), had its strongest abundance in the pellet fraction. This suggests that RNase Y might have precipitated or aggregated and was therefore found on the bottom of the gradient, potentially together with other components of the degradosome. While no direct proof, this result could be a starting point for the investigation of the existence of a degradosome complex in the pneumococcus.

As discussed in 1.2.5, Grad-seq not only allows the investigation of PPIs but also of RNPs. While the pneumococcus does not encode for any known global regulatory RBPs, it does contain universally conserved RNPs that are fundamental to the cell's functions. For example, the RNAP consisting of the α-, β-, β'- and ω-subunits (RpoA, RpoB, RpoC and RpoZ, respectively) co-migrated with 6S RNA (Figure 3.6 B), which was previously shown to bind RNAP in the Gram-positive *B. subtilis* (Burenina et al., 2014; Cavanagh et al., 2012; Trotochaud and Wassarman, 2005). The δ-subunit of RNAP, somewhat confusingly encoded by *rpoE*, is a subunit mostly found in low-GC Gram-positives such as Bacilli and Streptococci that is involved in regulating gene expression (Jones et al., 2003; Rabatinová et al., 2013; Xue et al., 2010). As expected, the δ-subunit was detected with the other core components of RNAP, indicating that it is a stable subunit in this essential complex.

Finally, the major σ-factor of the pneumococcus, σ^A (SigA), could also be detected together with the RNAP. The only other σ-factors encoded by the pneumococcus are the paralogous ComX1 and ComX2 (Tovpeko et al., 2016), which are exclusively expressed upon entering the competent state (see 1.3.2) and were therefore not detected in the dataset. Other major RNPs of the pneumococcus are represented by the SRP, the SmpB-tmRNA RNP, tRNA ligases and the ribosomal subunits, all of which showed excellent correlation in the gradient (Figure 3.6 B). Together, these results demonstrate that pneumococcus Grad-seq readily detects major stable protein-protein complexes as well as RNPs.
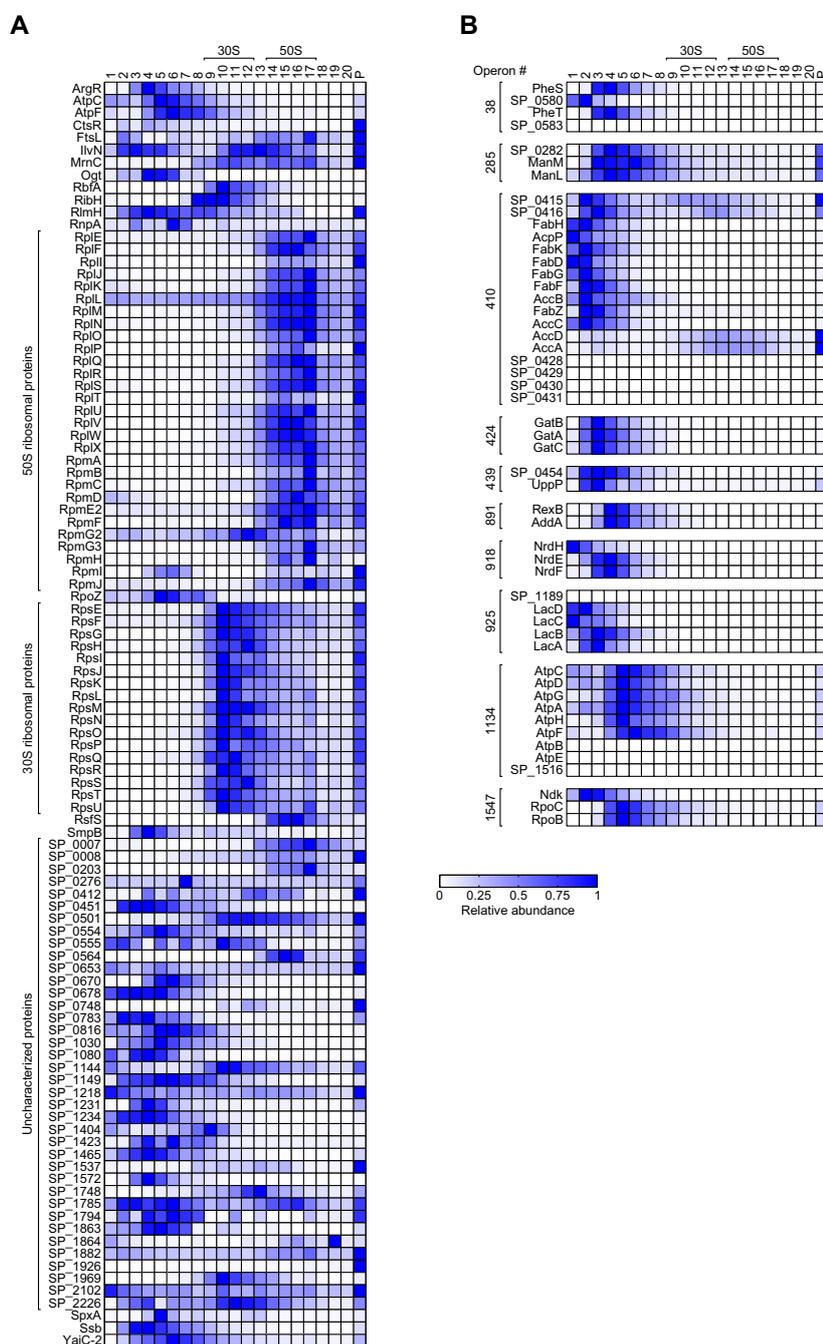
Figure 3.7: Grad-seq predicts complex formation. (**A**) Heat map showing the sedimentation profiles of 102 proteins with a molecular weight <20 kDa and peaks in fraction 4 or higher. Many ribosomal proteins are among these proteins as well as many proteins of unknown function. All of these proteins are predicted to form complexes. (**B**) Heat map showing the sedimentation profiles of selected operons that were predicted in a previous study (Warrier et al., 2018). Operons often encode for the subunits of a complex (Wells et al., 2016), making this approach a valuable resource for the prediction of functional interactions. Profiles are normalized to a maximum of 1.

3.3   Grad-seq assists in the characterization of protein functions

As shown for *E. coli* (see 2.9), Grad-seq can be used to predict whether a protein is likely involved in a complex. To test whether this assumption can be applied to the pneumococcus as well, the MS data were filtered for proteins <20 kDa that had their peak intensities in fraction 4 and above, resulting in a list of 102 proteins with these properties (Figure 3.7 A). As expected, 46 ribosomal protein were among these proteins. Ribosome-binding factor A (RbfA) and ribosomal silencing factor (RsfS) cosedimented with the 30S and 50S subunits, respectively, which agrees with previous findings that these proteins bind to non-assembled subunits (Häuser et al., 2012; Shajani et al., 2011). Further expected small proteins with peaks in high fraction numbers are RpoZ, SmpB and RnpA (the protein subunit of RNase P), all of which are involved in larger complexes and which were correctly identified by the applied filtering method.

The wealth of uncharacterized small proteins in this list of fast sedimenting proteins can be used as a starting point to predict their domain-based functions. For example, SP_1969 is annotated as a homolog of the 16S rRNA methyltransferase RsmD of *E. coli* (Lesnyak et al., 2007; Sergeeva et al., 2012). Indeed, SP_1969 cosedimented with the 30S subunit, suggesting that, like RsmD, it might be involved in 30S subunit maturation. The S4 domain-containing SP_0007 is a homolog of *E. coli* Hsp15, an RBP involved in 50S subunit recycling (Jiang et al., 2009). SP_0007 peaked with the 50S subunit, which suggests it might have the same function in the pneumococcus even though it does not contain the disordered ∼40 aa C-terminus of its *E. coli* homolog, which was shown to increase the affinity of Hsp15 to translationally inactive 50S subunits (Jiang et al., 2009). Another S4 domain-containing small protein, SP_2226, is a homolog of *B. subtilis* YaaA, which was suggested to be involved in 50S subunit assembly (Suzuki et al., 2014). In contrast to SP_1969 and SP_0007, this functional prediction did not fit with SP_2226: it coincided with the 30S subunit, implying it might have a different function in the pneumococcus.

A different way of using Grad-seq data is to predict protein interactions based on operon structures, which often encode the subunits of complexes on a single tran-

script (Wells et al., 2016). To test this assumption, operons as predicted by a recent study (Warrier et al., 2018) were analyzed for co-migrating proteins (Figure 3.7 B). This approach revealed that the acetyl-CoA carboxyltransferase complex (AccDA) also forms a complex in the pneumococcus and that, surprisingly, it mostly sedimented around fraction 13 and peaked in the pellet. A third peak was detected around fraction 3, congruent with biotin carboxylase (AccC) with which it forms a complex in *E. coli* (Cronan and Waldrop, 2002). Concerning the strong abundance of AccDA in the pellet, its *E. coli* homolog is known to act as a translational repressor of its own mRNA (Meades et al., 2010). If this is also true for pneumococcal AccDA, this might cause it to associate with polysomes, which for this big operon is especially intuitive since translatability of the other CDSs should not be affected by this interaction.

Another example is the RexB/AddA (a.k.a. RexAB) exonuclease/helicase involved in DNA repair in the pneumococcus (Halpern et al., 2004), which clearly co-migrated in the gradient, suggesting that it forms a stable complex. Finally, the membrane-associated, virulence-related protein UppP (a.k.a. BacA) (Chalker et al., 2000) surprisingly peaked around fraction 3, where SP_0454, a protein of unknown function with several trans-membrane domains, was also detected. While it is difficult to assess whether these membrane proteins remained properly folded during the Grad-seq protocol, their encoding in the same operon together with the observed co-sedimentation might indicate a functional interaction. Taken together, these results show that pneumococcal Grad-seq is able to assist in the identification and verification of functional PPIs.

3.4 Cbf1 is an sRNA-binding protein

3.4.1 MS2 pull-downs are not suitable for the pull-down of pneumococcal proteins

As mentioned before (see 1.3.2), a major goal of the present thesis was to identify novel RBPs in the pneumococcus. To pick potential candidate ncRNAs with similar gradient profiles, t-stochastic neighbor embedding (t-SNE) (van der Maaten and Hinton, 2008) analysis was performed in order to group ncRNAs according to their
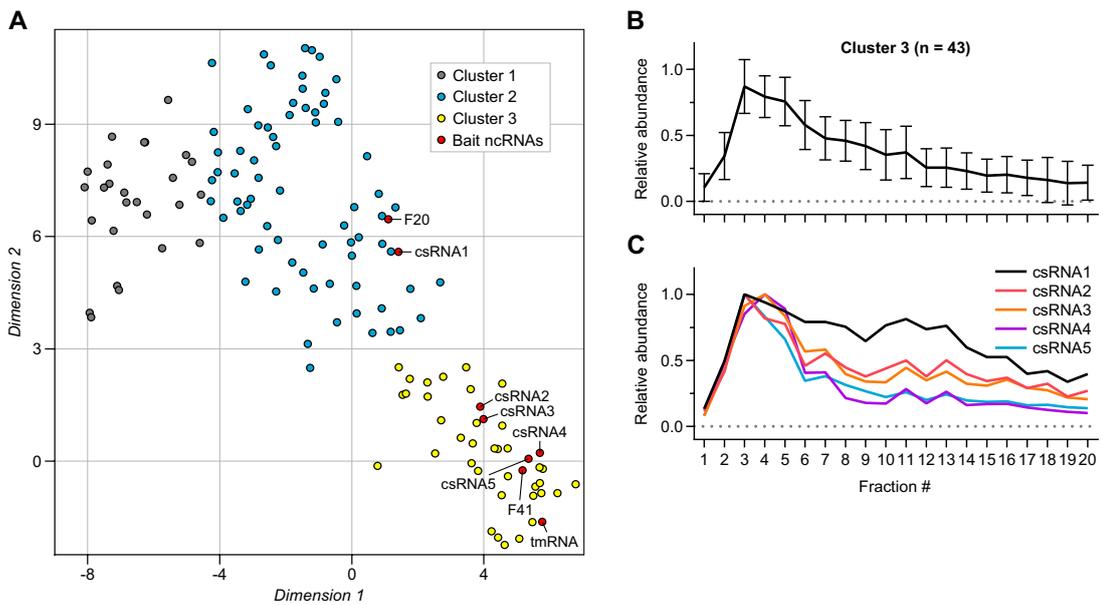
Figure 3.8: t-SNE analysis of the sedimentation of ncRNAs. (**A**) The sedimentation profiles of all ncRNAs is globally compared and visualized using t-SNE (van der Maaten and Hinton, 2008). The closer two ncRNAs (represented by dots) are in the plot, the more similar is their behavior in the gradient. The ncRNAs were further assigned into three different clusters (Cluster 1, 2 and 3) using k-means clustering (Lloyd, 1982). Cluster 3 contains several ncRNAs known to form RNPs, for example, tmRNA and M1 RNA, but also contains four of the five csRNAs. Based on this analysis, several ncRNAs (highlighted in red) were chosen to perform RNA-based pull-downs with. (**B**) Average sedimentation profile of the ncRNAs present in Cluster 3. The RNAs show an early peak in the beginning of the gradient, indicating involvement in small complexes. (**C**) The sedimentation profiles of the csRNAs show that, in the RNA-seq, csRNA1 has a slightly less pronounced peak in fraction 3 when compared to the other csRNAs. Profiles are normalized to a maximum of 1. Error bars show SD from the mean.
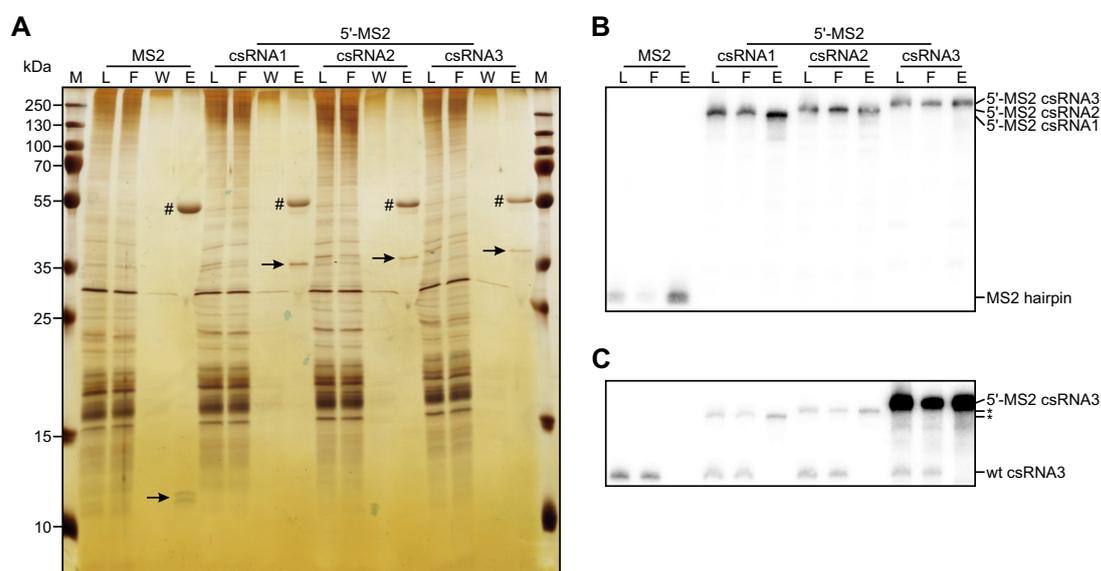
Figure 3.9: MS2 pull-downs using pneumococcal lysates. (**A**) Silver-stained gel of pull-downs performed using pneumcoccal lysates and 5′-MS2-tagged csRNA1, csRNA2 and csRNA3 as well as a control (MS2). Even though only two mild washes were performed, the second wash (W) contains no protein anymore. Similarly, no proteins can be detected in the eluate (E). The MS2-MBP (labeled with #) and the MS2-tagged RNAs (arrows) are recovered, however. M, size marker. (**B**) Northern blot control of the pull-down assay. The membrane was probed for the MS2 aptamer, revealing that both the control (MS2 hairpin) and the tagged csRNAs are recovered in the eluate. (**C**) The same membrane as in (**B**) was probed for csRNA3, verifying that wild-type (wt) csRNA3 cannot be recovered in the pull-down. Asterisks indicate remaining signals from the MS2 aptamer-specific probe. L, lysate (input control). F, flow-through.

migration in the gradient (Figure 3.8 A). As previously observed on northern blots (Figure 3.2 A), the csRNAs and F41 clustered together in the t-SNE plot (Figure 3.8 A), confirming their similar sedimentation more globally. Cluster 3, which they were part of, showed an average early peak in the beginning of the gradient, indicating involvement in small complexes (Figure 3.8 B). The only exception to this was csRNA1, which clustered a bit further away from the other csRNAs, probably due to the fact that, in the RNA-seq, its peak in fraction 3 was inexplicably slightly lower when compared to the other csRNAs (Figure 3.8 C). Still, all five csRNAs and F41 were chosen as promising candidates for pull-down assays to fish for interacting RBPs. Additionally, tmRNA and F20 were chosen as positive and negative controls, respectively, since tmRNA was known to bind SmpB (Keiler, 2015), whereas F20 is a riboswitch RNA that was not expected to specifically bind to proteins.

To perform the pull-downs, the RNAs were *in vitro* transcribed and tagged with a 5'- or 3'-located MS2 aptamer, which should allow capture of the RNAs and their potential binders by an immobilized MS2-MBP fusion protein (see 1.2.2.2). While successful in similar experiments in *Salmonella* and *E. coli* (Corcoran et al., 2012; Lalaouna et al., 2015; Said et al., 2009; Smirnov et al., 2016), this approach did neither allow pull-down of specific binders of the RNAs nor unspecific binders that might bind to, *e.g.*, the column resin (Figure 3.9 A). The second wash step already contained no detectable protein anymore and many attempts to optimize the experimental conditions resulted in no improvement. Still, pull-down of the MS2-MBP (labeled with #) and the tagged RNAs (arrows) did work as intended, which could further be verified via northern blotting (Figure 3.9 B). Probing for wild-type csRNA3 demonstrated that only the tagged RNA species could be recovered in the experiment, providing further evidence that the issue was not based on the pull-down of the RNAs (Figure 3.9 C).

In an attempt to increase the chance to pull-down specific RBPs bound to the csRNAs, glycerol gradients of pneumococcal lysates were run (Figure 3.10 A). As shown before (Figure 3.2 A), fractions 2–4 contain the highest amount of csRNAs, implying that a potential RBP should also be present in these fractions and be strongly enriched in comparison to the total lysate. Therefore, fractions 2–4 were used as the input for MS2 pull-downs as described above (Figure 3.10 A, highlighted in gray). The gradient-based "filtering" of the lysate drastically changed the protein composition of the input: the strong bands >100 kDa observed before (Figure 3.9 A) were almost fully depleted in this experiment (Figure 3.10 B). Yet again, the first wash removed most of the proteins that were present in the input, resulting in an empty second wash.

A slight improvement over the previous setup was that some unspecific proteins could be detected in the elution fraction, but only one band stood out to be enriched over the corresponding control: a faint band at around 12 kDa (blue arrows). MS analysis of this band did, however, not result in the identification of an enriched protein, indicating the observed band might have been an artifact from the silver staining. Importantly though, the general pull-down also worked with gradient fractions as the input, as indicated by the enrichment of MS2-MBP (labeled with
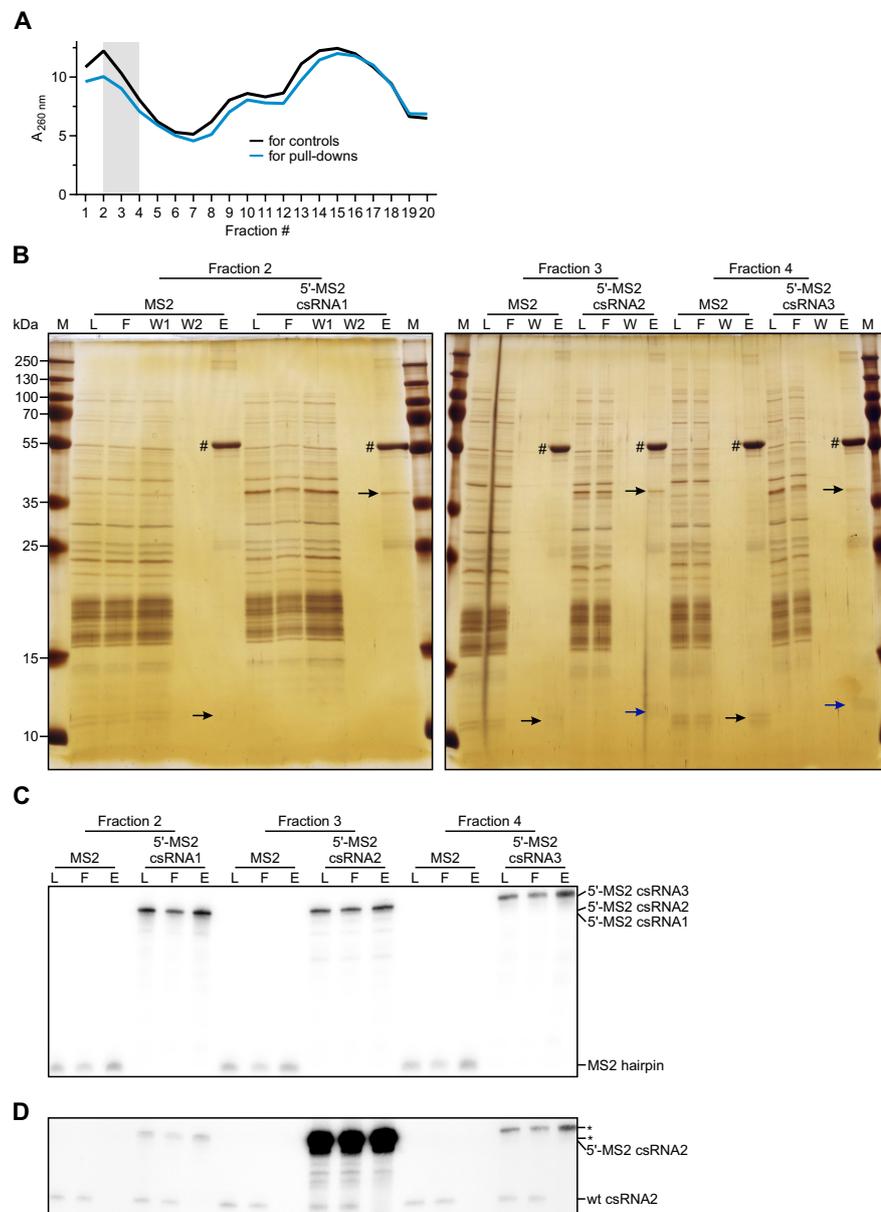
Figure 3.10: MS2 pull-downs using gradient fractions. (**A**) To enrich for interactors of the csRNAs, glycerol gradients were run using pneumococcal lysates. Fractions 2–4 (highlighted in gray) of these gradients were used to perform MS2 pull-downs. (**B**) Silver-stained gels of pull-downs performed using the gradient fractions shown in (**A**) and 5'-MS2-tagged csRNA1, csRNA2 and csRNA3 as well as a control (MS2). Even though only two mild washes were performed, the second wash (W2 for the left gel, W for the right gel) contains no protein anymore. The MS2-MBP (labeled with #) and the MS2-tagged RNAs (black arrows) are recovered, however. A faint band at ∼12 kDa is enriched in the pull-downs from fractions 3 and 4 (blue arrows) in comparison to the control, but MS analysis failed to identify any specific protein. W1, first wash. (**C**) Northern blot control of the pull-down assay. The membrane was probed for the MS2 aptamer, revealing that both the control (MS2 hairpin) and the tagged csRNAs are recovered in the eluate (E). (**D**) The same membrane as in (**C**) was probed for csRNA2, verifying that wild-type (wt) csRNA2 was present in the gradient fractions and that it cannot be recovered in the pull-down. Asterisks indicate remaining signals from the MS2 aptamer-specific probe. L, lysate (input control). F, flow-through.

#) and tagged RNAs (black arrows) in the eluates, which could further be verified by northern blotting (Figure 3.10 C). Finally, to test whether the selected gradient fractions indeed contained the wild-type csRNAs, which would be a pre-requisite for the pull-down of a specific RBP, northern blotting for csRNA2 was performed (Figure 3.10 D). This confirmed the presence of wild-type csRNA2 in all of the used fractions. In sum, these results demonstrate that MS2 pull-downs are not suitable for pneumococcal lysates.

3.4.2   A new pull-down approach reveals Cbf1 as an sRNA-binding protein

In the eukaryotic field, an *in vitro* pull-down approach based on a 14 nt long tag was successful in the identification of RBPs involved in microRNA biogenesis (Treiber et al., 2017). The concept of this "14mer pull-down" is similar to the MS2 approach: the RNAs of interest are *in vitro* transcribed using a DNA template that adds a 14 nt long sequence to the 5′ end of the RNA. Using this tag, the RNA can then be immobilized by binding to a complementary RNA adaptor that is built from 2′-O-methylated nucleotides and which carries a 3′ biotin, enabling the binding to streptavidin beads (Figure 3.11 A).

To test whether this setup works and whether the tagged RNA is stable during the protocol, the RNA was either coupled to the beads overnight or it was coupled overnight followed by overnight addition of a pneumococcal lysate. Following elution from the beads, an RNA gel was run and the recovered RNA visualized by ethidium bromide staining (Figure 3.11 B). Coupling of the tagged RNA to the beads worked, since it could readily be recovered after elution. More importantly, however, addition of pneumococcal lysate did not decrease the amount of recovered tagged RNA, demonstrating that the coupled RNA is stable over a long duration in lysate. It is noteworthy that pull-down of the tagged RNA from the lysate also recovered cellular RNA species (Figure 3.11 B, overexposed gel on the right), which could make this method a suitable alternative for the identification of RNA-RNA interactions as well.

Having established that the 14mer pull-down protocol should be suitable to capture RBPs from pneumococcal lysates, the ncRNAs highlighted in Figure 3.8 A were
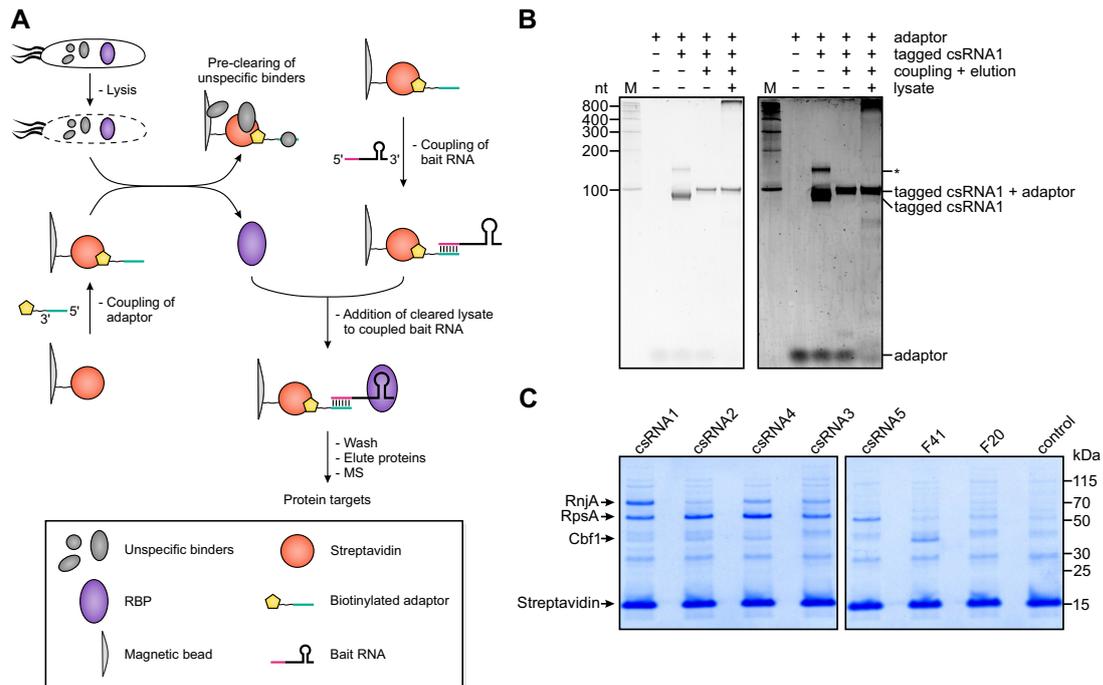
Figure 3.11: 14mer pull-downs reveal Cbf1 as an sRNA-interacting protein. (**A**) Overview of the 14mer pull-down protocol. After coupling of the adaptor to magnetic streptavidin beads, they are incubated with a lysate. Subsequently, the magnetic beads are removed, removing proteins that are unspecifically binding to the beads or the adaptor. The remaining, pre-cleared lysate is then incubated with streptavidin beads that are coupled to the adaptor and the *in vitro*-transcribed bait RNA. After washing, proteins specifically bound by the bait RNA are eluted and identified by MS. (**B**) RNA gel to test the coupling and stability of tagged csRNA1 using the 14mer pull-down protocol. After coupling to the beads and overnight incubation in buffer, tagged csRNA1 can be recovered from the beads. The same is true if an overnight incubation in pneumococcal lysate is added, revealing that the protocol is suitable to pull-down proteins from pneumococcal lysates. The gel on the right is an overexposed image of the gel on the left. The asterisk indicates a longer RNA species that was generated during the *in vitro* transcription of tagged csRNA1 and which was not recovered after coupling to the beads. M, size marker. (**C**) Coomassie-stained gels of pull-downs using several pneumococcal ncRNAs. Several proteins are enriched in comparison to the control, one of which is Cbf1. The riboswitch RNA F20 does not enrich Cbf1.
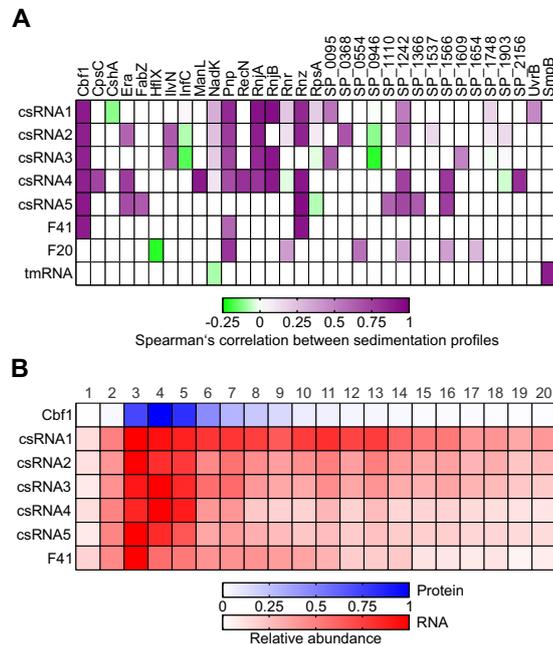
Figure 3.12: Cbf1 is a binder of the csRNAs. (**A**) Heat map showing Spearman's correlation coefficients between the sedimentation profiles of the bait ncRNAs used in the pull-downs and the proteins enriched by them (Figure 3.11 C). Ribosomal proteins except for S1 (RpsA) and proteins only co-purified with tmRNA (except for SmpB) are not shown. The sedimentation profile of Cbf1 shows strong correlation with those of the bait ncRNAs. (**B**) Heat map showing gradient profiles of Cbf1, csRNAs and F41. Cbf1 co-sediments with these sRNAs. Profiles are normalized to a maximum of 1.

tagged and subsequently used to perform pull-downs. An RNA oligonucleotide with the same sequence as the 14 nt tag served as the control. Excitingly, several specific protein bands were enriched for all bait RNAs when compared to the control (Figure 3.11 C). The only exception to this was the riboswitch RNA F20, which did not enrich any proteins to a degree that they were visible by Coomassie staining. In contrast to the MS2 pull-downs (see 3.4.1), unspecific background proteins were detected in all pull-downs, which can be used to improve the calculation of enrichment factors based on MS measurements.

Next, the Coomassie-stained gels were analyzed by MS in order to identify proteins that were enriched in the pull-down samples when compared to the control (Appendix Figure 7.5). As expected from the protein gels (Figure 3.11 C), several proteins could be identified this way. Figure 3.12 A shows all proteins that were pulled down by the respective RNAs with the following exceptions: Except for ribosomal protein S1 (RpsA), all other ribosomal proteins were omitted for clarity.

Similarly, all proteins (except for SmpB) that were enriched only by tmRNA were omitted. Interestingly, most of the degradosome subunits, namely CshA, PNPase and RNases J1 and J2, were enriched by at least one of the bait RNAs. Correlation analyses between the sedimentation profiles of the bait RNAs and the enriched proteins showed, however, that only RNases J1 and J2 correlated well. The essential RNase Z, which is involved in tRNA maturation (Durand and Condon, 2018), was pulled down by F41 as well as all csRNAs except for csRNA3 and its gradient profile strongly correlated with those of these RNAs.

However, only Cbf1 was enriched by all csRNAs and F41 and additionally showed excellent correlation within the gradient (Figure 3.12 A and B). Importantly, the negative control riboswitch RNA F20 did not enrich most of the proteins pulled down with the csRNAs or F41 (Figure 3.12 A). Similarly, the positive control RNA tmRNA only enriched one protein that was detected by other baits (the NAD kinase NadK) and was the only one to specifically enrich its physiological target SmpB (Keiler, 2015). These results show that the new 14mer pull-down approach allows the capture of specific binders of tagged pneumococcal ncRNAs and that Cbf1 is an interacting protein of the csRNAs and F41.

3.4.3 Function and conservation of Cbf1 and its homologs

Cbf1 (*cmp*-binding factor 1) was first described as a host factor for replication of plasmid pT181 in *Staphylococcus aureus* (Zhang et al., 1997). Following this first study, *S. aureus* Cbf1 and its *B. subtilis* homolog YhaM were shown to be manganese-dependent 3′→5′ exonucleases whose activity was blocked by stem-loop structures in their RNA substrates (Oussenko et al., 2002). YhaM was further shown to be able to degrade DNA, potentially even with higher activity than RNA (Fang et al., 2009; Oussenko et al., 2002) and implicated in mRNA turnover (Oussenko et al., 2005) as well as 23S rRNA maturation (Redko and Condon, 2010). In *Streptococcus pyogenes*, YhaM was recently shown to only trim its targets by nibbling away a few nucleotides from the 3′ end until it reached a stem-loop structure like a ρ-independent terminator (Lécrivain et al., 2018), thereby agreeing with the previous observation in *B. subtilis* (Oussenko et al., 2002). Yet, the specific physiological functions of these proteins remain unknown.
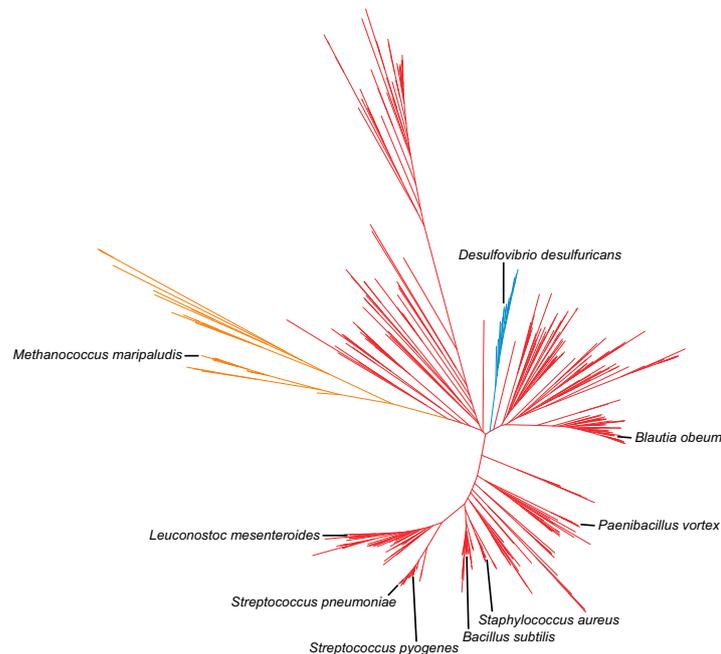
Figure 3.13: Phylogenetic analysis of Cbf1. Phylogenetic analysis of Cbf1 based on 508 protein sequences deposited in eggNOG 4.5.1 (Huerta-Cepas et al., 2016) (COG3481). Cbf1 homologs were identified in archaea (orange), Firmicutes (red) and δ-proteobacteria (blue).

Cbf1/YhaM homologs are present in a wide range of Gram-positives and a minority of Gram-negatives as well as archaea (Figure 3.13), implying that their functions are generally either absent in Gram-negative bacteria or fulfilled by a different enzyme. The HD domain present at the C-terminus of Cbf1/YhaM proteins likely is the catalytic domain (Oussenko et al., 2002). It is named after the conserved histidine (H) and aspartate (D) residues that are found within it and often is involved in enzymes involved in nucleic acid metabolism (Aravind and Koonin, 1998). Prominent examples of HD domain-containing proteins are RNase Y, SpoT and RelA. In the endoribonuclease RNase Y, the HD domain is the catalytically active domain (Durand and Condon, 2018). The bifunctional SpoT, which is involved in the stringent response of *E. coli*, has weak ppGpp synthetic activity and strong ppGpp degrading activity, the latter of which is carried out by its HD domain (Atkinson et al., 2011). RelA shares the same ancestral homolog as SpoT but its HD domain is inactive, explaining why RelA has strong ppGpp synthetic activity and no ppGpp degrading activity (Atkinson et al., 2011). The conserved H and D residues of HD domains

are also conserved in Cbf1/YhaM proteins (Appendix Figure 7.6) and are therefore likely involved in the catalytic activity.

The other domain of Cbf1/YhaM homologs is a tRNA_anti domain, which is a family of OB-fold-like (OB = oligonucleotide/oligosaccharide-binding) domains found in some aminoacyl-tRNA synthetases (El-Gebali et al., 2019). OB-fold domains are usually involved in the RNA or DNA binding of the proteins they are present in (Flynn and Zou, 2010; Theobald et al., 2003). Many well-known classes of RBPs contain OB-fold or OB-fold-like domains, including cold shock proteins, the ρ transcription termination factor and several ribosomal proteins (Mitchell et al., 2019). The OB-fold-like domain of Cbf1/YhaM proteins is located at the N-terminus and is likely involved in their RNA-binding properties (Appendix Figure 7.6).

3.4.4   Validation of the sRNA-binding activity of Cbf1

To validate the initial findings of the sRNA-binding activity of Cbf1 (see 3.4.2), the reverse experiment was performed by *in vivo* CLIP-seq (see 1.2.2.1). To allow purification of Cbf1, a *cbf1*-3xFLAG strain was used. After crosslinking, immunoprecipitation, gel electrophoresis and membrane transfer, a ∼5–10-fold enrichment of RNA compared to the non-crosslinked control was observed, demonstrating that RNA can indeed be specifically crosslinked to Cbf1 (Figure 3.14 A). Following the release of the RNA fragments bound to Cbf1, RNA-seq was performed in order to assign the fragments to their respective transcripts. This resulted in 528 statistically significant peaks ($\log_2$ fold change >1, adjusted p-value <0.01), 354 (∼67 %) of which mapped to CDSs (Figure 3.14 B). 94 (∼18 %) peaks derived from intergenic sequences, which originated from unannotated antisense transcripts, intergenic transcripts and UTRs. 38 (∼7 %) peaks could be attributed to 30 unique annotated ncRNAs (Figure 3.14 B and C).

F41 was the most enriched ncRNA in the CLIP-seq experiment, closely followed by csRNA1, csRNA3 and csRNA5 (Figure 3.14 C). Unexpectedly, csRNA2 and csRNA4 were absent from the list of enriched transcripts, which might have been caused by unfavorable positioning of crosslinkable nucleotides in the Cbf1-RNA complex. The abundant house-keeping ncRNAs 6S RNA, tmRNA and M1 RNA tend to crosslink to RBPs in CLIP-seq experiments (Holmqvist et al., 2018) and were also
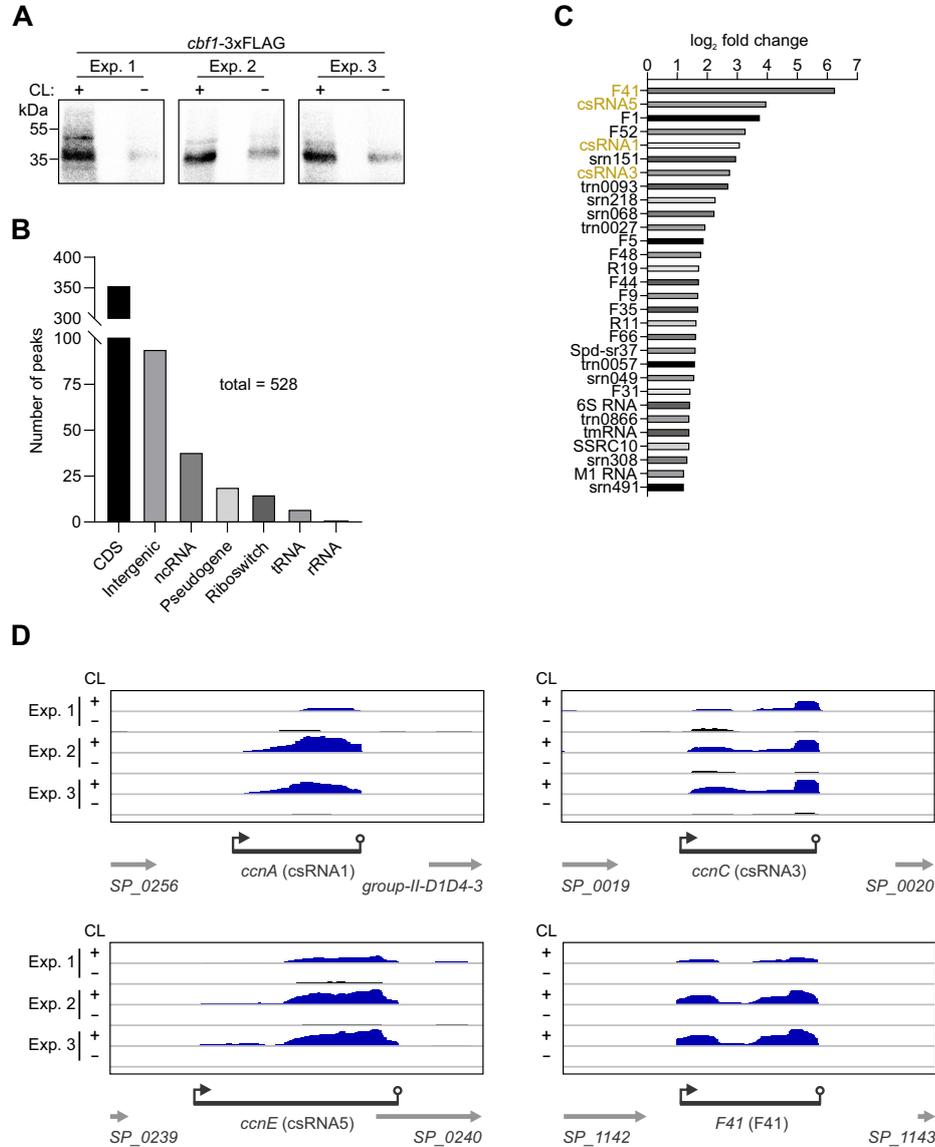
Figure 3.14: CLIP-seq of Cbf1. (**A**) Autoradiographs of radioactively labeled RNA fragments covalently bound to Cbf1 after *in vivo* UV crosslinking, immunoprecipitation, gel electrophoresis and membrane transfer. After purification of the RNA, sequencing was performed. CL, UV crosslinking. Exp., experiment. (**B**) Cbf1 CLIP-seq identified 528 peaks, most of which derived from coding sequences (CDS). (**C**) Enrichment of the 30 unique ncRNAs identified by Cbf1 CLIP-seq. F41, csRNA1, csRNA3 and csRNA5 were previously used to pull-down Cbf1 (Figure 3.11 C) and are strongly enriched in the CLIP-seq data (highlighted in orange). (**D**) CLIP-seq coverage plots of F41, csRNA1, csRNA3 and csRNA5 reveal enrichment of reads at the 3' ends of these transcripts, which is dependent on crosslinking.

Figure 3.15: Cbf1 trims its targets. (**A**) Northern blotting of *in vitro* RNase assays using recombinant Cbf1 and *in vitro*-transcribed RNAs. Cbf1 cleaves csRNA1, csRNA3 and csRNA5 in a manganese-dependent fashion, whereas it has no activity on 5S rRNA. (**B**) The same experiment as in (**A**) was performed using truncated csRNA1, csRNA3 and csRNA5, all of which were missing the single-stranded uridine stretch following the ρ-independent terminator at their 3′ ends. Cbf1 is not able to cleave these RNAs. Wild-type (wt) RNA species used in (**A**) were loaded as reference. (**C**) Northern blotting of total RNA obtained from wt, Δ*cbf1* and complementation strains shows that Cbf1-dependent cleavage of the csRNAs also occurs *in vivo*. 5S rRNA was used as loading control and is not affected by knockout of *cbf1*.

slightly enriched in the present dataset. The obtained reads of the CLIP-seq experiment mostly mapped to the 3′ ends of the transcripts, whereas little coverage was observed around the 5′ ends (Figure 3.14 D), supporting the predicted role of Cbf1 as a 3′→5′ exonuclease (see 3.4.3). These results show that Cbf1 binds a large variety of transcripts *in vivo*, among which the csRNAs and F41 could be verified.

3.5 Functional characterization of Cbf1

3.5.1 Cbf1 is a manganese-dependent 3′→5′ exonuclease

To test whether pneumococcal Cbf1 is a manganese-dependent 3′→5′ exonuclease similar to its previously studied homologs in *B. subtilis*, *S. aureus* and *S. pyogenes* (Lécrivain et al., 2018; Oussenko et al., 2002), recombinant Cbf1 was purified from *E. coli* and incubated with *in vitro*-transcribed RNAs (Figure 3.15 A). In the presence of manganese ions, immediately after addition of Cbf1 to the reaction, a slight shift toward a faster migrating species of RNA was detectable by northern blotting in the

case of csRNA1 and csRNA3 (time point 0 min). After 5 and 15 min of incubation at 37 °C, this shift was more noticeable and also visible for csRNA5. Since the shortened bands were not accompanied by other species and they were of similar intensity than the undigested bands, this suggested complete digestion of the original RNAs. The observed shortening was also fully dependent on Cbf1 and manganese ions, thus verifying the predictions based on *B. subtilis* YhaM (Fang et al., 2009; Oussenko et al., 2002). *In vitro*-transcribed 5S rRNA, which *in vivo* is processed by RNase M5 (Bechhofer and Deutscher, 2019), could not be processed by Cbf1.

To assess whether cleavage by Cbf1 indeed happens at the single-stranded overhang that follows the ρ-independent terminator of the tested transcripts, truncated versions of csRNA1, csRNA3 and csRNA5 missing the 3′ U-stretch were *in vitro* synthesized and subjected to the same experiment as before. Intriguingly, no decay could be observed in this case, implying that Cbf1 indeed only cuts off a few nucleotides before being stopped by a stem-loop structure (Figure 3.15 B). Having established that Cbf1 shortens csRNAs *in vitro*, a *cbf1* knockout mutant and the respective complementation strain were constructed to allow *in vivo* validation. Total RNA extraction followed by northern blotting revealed what was expected from the *in vitro* experiments: the tested RNAs were longer in the knockout strain than in the wild type (Figure 3.15 C). Again, processing of 5S rRNA was not affected and no difference between the wild type and the mutant could be observed. Taken together, these results show that pneumococcal Cbf1 is a manganese-dependent 3′→5′ exonuclease that only trims its targets.

3.5.2   Cbf1 stabilizes its target ncRNAs

As mentioned before (see 3.4.3), the exact physiological functions of Cbf1/YhaM proteins is not well understood. For example, in *S. pyogenes*, knockout of *yhaM* has no global effect on RNA levels, even though it causes the same phenotype as observed here for *S. pneumoniae, i.e.*, it shortens its targets (Lécrivain et al., 2018). To investigate what the repercussions of this trimming could be, rifampicin RNA stability assays were performed comparing wild-type pneumococcus to a *cbf1* knockout strain.

The antibiotic rifampicin shuts down the cellular transcription machinery and thereby allows monitoring of RNA decay over time. Unexpectedly, knockout of *cbf1*
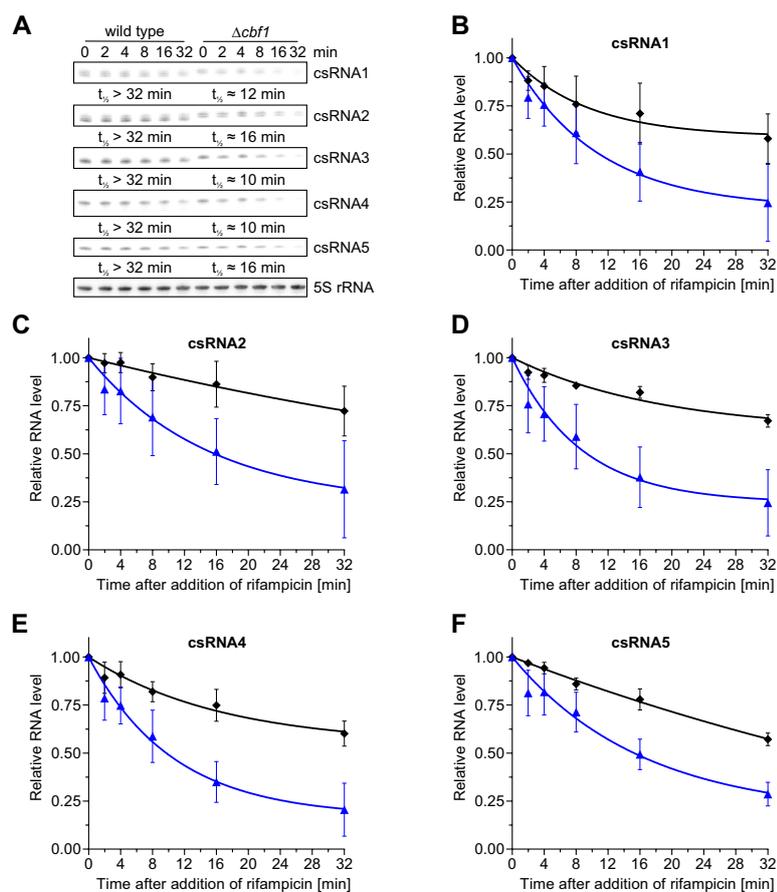
Figure 3.16: Cbf1 stabilizes csRNAs *in vivo*. (**A**) Rifampicin *in vivo* RNA stability assays show an increased stability of csRNAs upon deletion of *cbf1*. RNA half-lives were calculated based on the quantifications shown in (**B–F**). Error bars show SD from the mean, n ≥ 2.

caused a strong decrease in stability for all csRNAs compared to the wild type (Figure 3.16 A), lowering their half-lives from >32 min to ∼10–16 min (Figure 3.16 B–F). This indicated a protective function of Cbf1 for these transcripts. While classic RNA chaperones such as Hfq are well-known to increase the stability of their targets (Holmqvist and Vogel, 2018), this function is generally not expected for RNases such as Cbf1. Only one other example of RNase-dependent transcript stabilization has been previously described in bacteria: sRNA stabilization by PNPase (Cameron et al., 2018, 2019). How Cbf1 might cause this effect will be discussed separately (see 3.6).

3.5.3   *cbf1* is part of the competence regulon

Binding and trimming of csRNAs by Cbf1 putatively connected *cbf1* to the competence regulon, which is one of the fundamental systems of pneumococcal virulence (Salvadori et al., 2019). As introduced in 1.3.2, csRNAs post-transcriptionally downregulate *comC* (Laux et al., 2015; Schnorpfeil et al., 2013), which codes for the precursor peptide of CSP, which in turn is the pheromone that triggers the activation of the competence regulon (Shanker and Federle, 2017). Apart from its putative implication in the competence regulon due its direct interaction with csRNAs (see 3.4.4), global studies also predicted *cbf1* to be a late competence gene upregulated by the alternative paralogous σ-factors ComX1 and ComX2 (Peterson et al., 2004; Slager et al., 2019).

To verify this prediction, a *cbf1*-3xFLAG culture was stimulated with CSP and sampled over time, extracting both protein and RNA (Figure 3.17 A). Western blotting revealed that Cbf1 protein levels indeed were upregulated ∼2-fold 15 min after stimulation with CSP (Figure 3.17 B), which is the expected induction speed of a late competence gene (Shanker and Federle, 2017). GAPDH levels on the other hand were not influenced by the pheromone (Figure 3.17 A). As expected, the early competence gene operon *comCDE* (Shanker and Federle, 2017) was strongly upregulated on RNA level 5 min after CSP addition, thereby confirming that the stimulation worked as intended. It is of note that northern blotting revealed the *comCDE* polycistronic mRNA to be heavily processed.
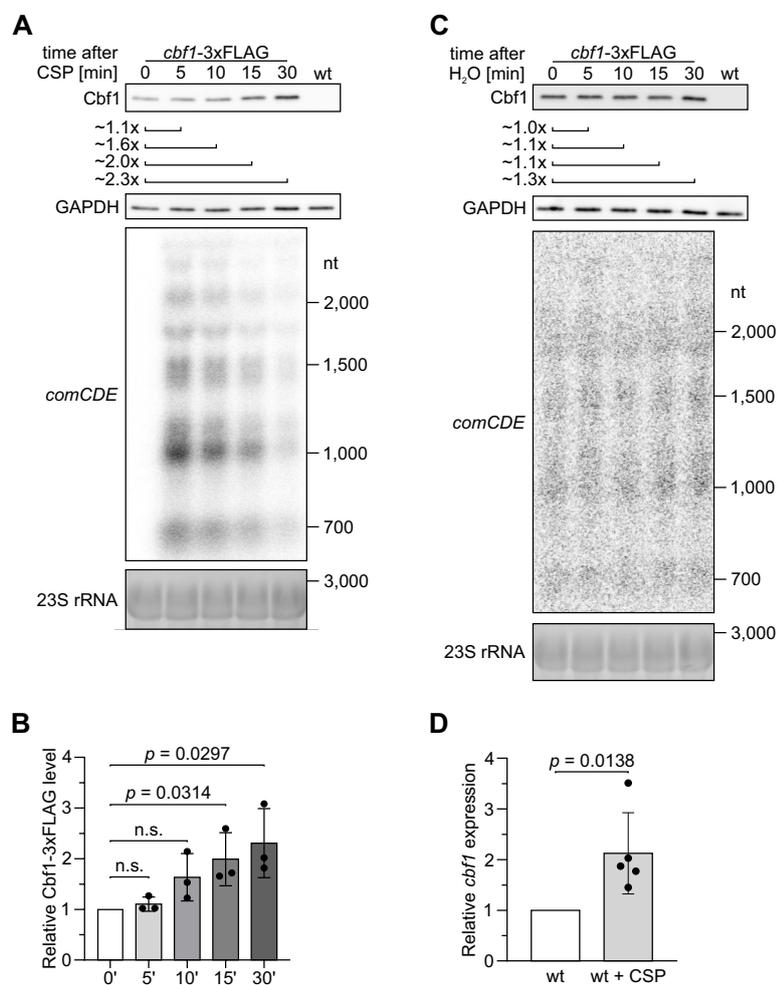
Figure 3.17: *cbf1* **is a member of the competence regulon.** (**A**) Western blotting of Cbf1-3xFLAG reveals that its levels increase ∼2-fold 15 min after induction with competence stimulating peptide (CSP). Agarose northern blotting indicates that the early competence gene operon *comCDE* is upregulated 5 min after induction and served as a control for the experiment. GAPDH and 23S rRNA served as loading controls. (**B**) Quantification of the western blot results shown in (**A**). Error bars show SD from the mean, n = 3. *p*-values were calculated using an unpaired, two-tailed t-test. (**C**) The same experiment as in (**A**) was performed using a mock induction with water instead of CSP. Neither Cbf1 nor *comCDE* show upregulation. GAPDH and 23S rRNA served as loading controls. (**D**) RT-qPCR of *cbf1* transcript levels 10 min after induction with CSP shows an upregulation of ∼2-fold compared to the non-induced control. Error bars show SD from the mean, n = 5. *p*-values were calculated using an unpaired, two-tailed t-test.
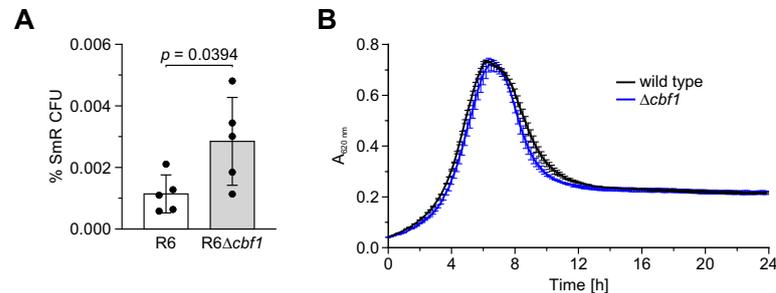
Figure 3.18: Cbf1 negatively regulates competence. (**A**) Spontaneous competence assays comparing wild type and Δ*cbf1* strains in the R6 background show a ∼2.5-fold increase in obtained streptomycin-resistant transformants in the knockout. This indicates that Cbf1 is a negative regulator of competence. Error bars show SD from the mean, n = 5. *p*-values were calculated using an unpaired, two-tailed t-test. (**B**) Growth curves in rich medium comparing wild-type and Δ*cbf1* strains do not reveal a growth defect of the knockut.

To exclude that the observed upregulation of *cbf1* was caused by continuous growth of the culture, the same experiment was repeated applying a mock induction with water instead of CSP (Figure 3.17 C). In this case, only a minor increase in Cbf1 protein levels was observed, whereas GAPDH levels again did not change over the course of the experiment. Control of the experiment was again performed by northern blotting of total RNA extracted at the indicated time points, revealing no upregulation of the *comCDE* operon after mock stimulation with water. Interestingly, even though only a weak signal could be detected, the same processing pattern of the mRNA could be observed in this case when compared to the CSP-stimulated experiment (Figure 3.17 A). Finally, *cbf1* mRNA levels were measured 10 min after CSP stimulation using RT-qPCR, revealing a ∼2-fold increase in RNA levels (Figure 3.17 D). Taken together, these results establish *cbf1* as a member of the late competence genes of the pneumococcus.

3.5.4   Cbf1 is a negative regulator of competence

The identified stabilization of csRNAs by Cbf1 and its upregulation during competence suggested that Cbf1 might negatively regulate competence. To test this hypothesis, spontaneous competence assays were performed that are independent of exogenously added CSP. This is important since the csRNAs post-transcriptionally downregulate expression of CSP (Laux et al., 2015; Schnorpfeil et al., 2013), meaning that exogenous CSP would probably mask the effect of the csRNAs. Competence in-

duction was surveyed by addition of a PCR product of *rpsL* (ribosomal protein S12) containing a point mutation conferring resistance against streptomycin (Muschiol et al., 2017). This allowed quantification of the competence assays by counting colony forming units (CFUs) after selection on streptomycin-containing plates.

However, no transformants were obtained using the TIGR4 strain background. To bypass this nuisance, strain R6, which is known to spontaneously develop competence when grown at pH 7.9–8.0 (Moscoso and Claverys, 2004), was used instead of TIGR4. Intriguingly, knockout of *cbf1* increased the number of CFUs in the R6 background by ∼2.5-fold when compared to the wild type (Figure 3.18 A), thereby suggesting Cbf1 to indeed be a negative regulator of competence. The *cbf1* deletion mutant did not exhibit any difference in growth when compared to the wild type (Figure 3.18 B), showing that this was not the cause of the observed competence phenotype. These results suggest that Cbf1 is negative regulator of competence.

3.6 Discussion

The overwhelming majority of studies on sRNAs as well as their targets and mechanisms were performed in the closely related Gram-negative Enterobacteria *E. coli* and *Salmonella* (Hör et al., 2020b). Results from these studies suggest that sRNAs generally need an RNA chaperone like Hfq in order to fulfill their functions, which most of the time is post-transcriptional regulation of one or more mRNA targets (Gorski et al., 2017; Holmqvist and Vogel, 2018). In Gram-positive organisms, however, much less is known about sRNA mechanisms and the potential RBPs involved. Hfq, for instance, is only present in some Gram-positives and its function in these organisms is poorly understood.

In *B. subtilis*, Hfq seems to only play a minor role in gene regulation, even though several groups have tried to identify its functions (Dambach et al., 2013; Hämmerle et al., 2014; Rochat et al., 2015). Similarly, in *Listeria monocytogenes*, Hfq is not involved in general regulation by sRNAs but instead was shown to facilitate antisense regulation (Nielsen et al., 2010). In contrast, the function of Hfq in the nosocomial pathogen *Clostridium difficile* is potentially more similar to its enterobacterial homolog: *C. difficile* Hfq can partially replace Hfq in *E. coli* (Caillet et al.,

2014) and its depletion has pleiotropic effects (Boudry et al., 2014). Nonetheless, Gram-positives make ample use of sRNA-mitigated regulation (Brantl and Brückner, 2014; Desgranges et al., 2019; Quereda and Cossart, 2017; Wagner and Romby, 2015; Wassarman, 2018), raising the question whether there might be overlooked regulatory RBPs in these organisms.

Similar to the examples mentioned above, the clinically important pneumococcus transcribes a vast number of ncRNAs, some of which were shown to be important for several physiological processes and virulence (Laux et al., 2015; Mann et al., 2012). However, the functions and molecular mechanisms of most of them remain unknown. The pneumococcus further encodes no homologs of the well-studied major regulatory RBPs CsrA, Hfq or ProQ (Tettelin et al., 2001). Therefore, Grad-seq was used here in order to provide an RNA/protein complexome dataset for the pneumococcus and with it the first of its kind for a Gram-positive organism. Importantly, apart from the lysis conditions, no adjustments to the Grad-seq protocol established with *E. coli* in the previous chapter had to be performed in order to obtain suitable separation of the pneumococcal lysate (Figure 3.1). This simultaneously suggests that the Grad-seq protocol may be generally applicable to bacteria without major changes, making it a robust choice for the analysis of bacterial complexomes.

The only previous global study investigating the protein interactome of the pneumococcus was based on Y2H (Wuchty et al., 2017), which is limited in several ways since it relies on binary interactions that have to be established in a different species (see 1.2.1). In contrast, Grad-seq is independent of tagging, allowing to draft a global landscape of stable complexes from a single experiment (Smirnov et al., 2017a). With this, Grad-seq for the first time revealed major pneumococcal complexes such as the ATP synthase F_1 complex (Figure 3.6 A). Moreover, the combined analysis of the global RNA and protein datasets here provided an overview of conserved stable RNPs such as 6S RNA-RNAP or the SRP, demonstrating that these universally conserved complexes can be readily reproduced by Grad-seq (Figure 3.6 B).

In comparison to the binary Y2H approach (Wuchty et al., 2017), Grad-seq does not provide predictions of direct interactions. Yet, sedimentation properties of proteins can be cross-referenced with functional predictions in order to evaluate the likeliness of the predictions. Here, SP_1969 was detected to co-migrate with the 30S

subunit (Figure 3.7 A). Based on database information from InterPro (Mitchell et al., 2019), SP_1969 belongs to a family of 16S rRNA methyltransferases, which *E. coli* RsmD is also part of. RsmD methylates G966 of the 16S rRNA (Lesnyak et al., 2007), which stabilizes the binding of initiator tRNA to the 30S pre-initiation complex and whose absence causes reduced fitness when compared to the wild type (Burakovsky et al., 2012).

Purified *E. coli* RsmD was further shown to tightly bind to non-methylated 30S subunits after *in vitro* reconstitution and sucrose gradient analysis (Sergeeva et al., 2012). It therefore seems likely that SP_1969 indeed is a homolog of RsmD. Unexpectedly, in the *E. coli* Grad-seq presented in this thesis (see 2), RsmD did not co-migrate with the 30S subunit but rather exhibited slower sedimentation with a peak around fraction 5. One reason could be that, under the specific conditions used here, the *E. coli* RsmD-30S subunit interaction was not stable enough to survive the protocol.

In contrast to SP_1969, the *E. coli* Hsp15 homolog SP_0007 co-sedimented with the 50S subunit with almost no intensity in the pellet fraction (Figure 3.7 A). *E. coli* Hsp15 is involved in the recycling of dead-end 50S subunits, which have stalled polypeptide chains attached and are no longer able to form new translating ribosomes (Jiang et al., 2009; Korber et al., 2000). The observed sedimentation of SP_0007 fits this functional prediction, suggesting it might indeed be a functional homolog of Hsp15. Further evidence for this comes from the *E. coli* Grad-seq dataset, where Hsp15 exclusively co-migrated with the 50S subunit too (Figure 2.23).

In comparison to the two examples above, Grad-seq is also able to challenge functional predictions: *B. subtilis* YaaA was suggested as a protein involved in 50S subunit assembly (Suzuki et al., 2014). In the present dataset, the pneumococcal homolog of YaaA, SP_2226, co-sedimented with the 30S subunit instead, thereby questioning functional conservation between the two species (Figure 3.7 A). Interestingly, YaaA homologs within the family of Streptococcaceae have a ~50 aa long extended N-terminus not present in other families of Gram-positives (Appendix Figure 7.7). Given that *B. subtilis* YaaA is only 71 aa in length and SP_2226 is 122 aa, it is easy to imagine that this extension could have a drastic influence on the protein's function.

Apart from refining functional predictions of uncharacterized proteins, the re-
source presented here also enables the search for biochemically similar classes of
RNAs, which are defined solely by their sedimentation behavior in the gradient,
independent of sequence or length of the RNAs included in them (Smirnov et al.,
2017a). The clustering of ncRNAs in this chapter revealed three major classes of
ncRNAs in the pneumococcus, one of which contained several ncRNAs known to
be involved in RNPs as well as 4 of the 5 csRNAs and the functionally unknown
F41 (Figure 3.8 A). Based on this, pull-down assays using a newly developed pull-
down approach inspired by a similar protocol used for eukaryotes (Treiber et al.,
2017, 2018) was employed in order to identify a potential common interactor of
the csRNAs (Figure 3.11). Indeed, the 3′→5′ exonuclease Cbf1 turned out to be a
common interactor using this pull-down approach, which subsequently was ver-
ified using CLIP-seq (Figure 3.14). These data further revealed that Cbf1 targets
many different transcripts, especially mRNAs. Similarly, *S. pyogenes* YhaM trims the
majority of 3′ ends of mRNAs generated via either a transcription termination or
endonucleolytic cleavage event (Lécrivain et al., 2018). However, the consequences
of this global targeting are not understood.

Another result from the pull-down experiments is that the pneumococcus pos-
sibly does not encode for a global RBP like Hfq or ProQ. While the existence of
such a protein cannot fully be excluded at this point, the existence of the more
specialized KhpA/B complex (Figure 3.5), which binds subsets of sRNAs, tRNAs
and mRNAs (Zheng et al., 2017), suggests that the pneumococcus and similar or-
ganisms may either rely on such narrow spectrum RBPs or that their RNA-based
regulation at least partially occurs without the aid of a chaperone. To elucidate this
in more detail, alternative orthologous strategies for the identification of RBPs will
be necessary in the future. A recently introduced method relies on the isolation of
crosslinked RBP-RNA complexes via organic phase extraction or adhesion to sil-
ica matrices and is suitable for bacteria, providing a promising platform for the
discovery of RBPs in the pneumococcus (Queiroz et al., 2019; Shchepachev et al.,
2019; Urdaneta et al., 2019). Moreover, the gradient-based R-DeeP (RNA-dependent
proteins) is able to identify RNA-dependent proteins by comparing the shifting of

proteins between an RNase-treated gradient and a control gradient and is applicable to bacteria (Caudron-Herger et al., 2019, 2020).

Similar to its homologs in *B. subtilis*, *S. aureus* and *S. pyogenes* (Fang et al., 2009; Lécrivain et al., 2018; Oussenko et al., 2002, 2005), pneumococcal Cbf1 was shown *in vitro* to trim the csRNAs in a manganese-dependent fashion; a reaction that was blocked by stem loop structures (Figure 3.15 A). This trimming also occurred *in vivo*, thus recapitulating the *in vitro* observations with purified Cbf1 (Figure 3.15 C). Unexpectedly, this shortening by Cbf1 increased the stability of the csRNAs *in vivo* (Figure 3.16), which downregulate competence by inhibiting translation of *comC* (Laux et al., 2015). In a knockout of *yhaM*, *B. subtilis* RNase R was observed to be more active than in presence of *yhaM*, implying a protective function of YhaM (Oussenko et al., 2005). Given that RNase R has helicase activity, enabling it to continue decay in the presence of stem-loops (Bechhofer and Deutscher, 2019), removal of single-stranded 3' regions by Cbf1/YhaM might remove the foothold RNase R needs in order to bind its targets.

Indeed, in *B. subtilis*, a 12 nt long single-stranded region was not enough for RNase R to allow binding (Oussenko and Bechhofer, 2000), which could explain the protective function of Cbf1/YhaM. Another possible explanation is that Cbf1 does not dissociate after trimming of its targets and thereby shields them from other RNases, similarly to how ProQ protects its targets from degradation by RNase II in *Salmonella* (Holmqvist et al., 2018) and by PNPase in *Neisseria* (Bauriedl et al., 2020). The only other bacterial exonuclease known to exert a protective effect so far is PNPase in Enterobacteria, which degrades certain sRNAs and protects others, a function dependent on Hfq (Andrade et al., 2012; Bandyra et al., 2016; Cameron and De Lay, 2016; Cameron et al., 2018, 2019; De Lay and Gottesman, 2011). In eukaryotes, this phenomenon was observed for several classes of ncRNAs that can be stabilized by PARN- or TOE1-dependent trimming at their 3' ends (Berndt et al., 2012; Son et al., 2018; Tseng et al., 2015).

The direct interaction between Cbf1 and csRNAs (see 3.4.4) raises the question whether Cbf1 is playing a role in the competence regulon. Intriguingly, others have suggested *cbf1* to be a member of ComX-induced late competence genes (Peterson et al., 2004; Slager et al., 2019), which are upregulated approximately 12.5–15 min

after induction by CSP (Shanker and Federle, 2017). Proving their prediction right, *cbf1* could be shown to be upregulated on both RNA and protein level 15 min after induction (Figure 3.17). As an exonuclease, Cbf1 is unlikely to contribute to the major competence complexes needed, *e.g.*, for the uptake or integration of DNA. However, since Cbf1 stabilizes the csRNAs (Figure 3.16), Cbf1 in return should have a negative effect on competence too. Supporting this prediction, a knockout strain of *cbf1* was found to be more competent when compared to the wild type (Figure 3.18 A).

Still, as shown here by CLIP-seq, Cbf1 targets many more transcripts in the cell than just the csRNAs (Figure 3.14 B), suggesting there might be additional transcripts involved in the observed phenotype. Further, in a different strain of *S. pneumoniae*, another study showed reduced (20 % down) competence in a *cbf1* knockout background (Peterson et al., 2004). However, in this case, exogenous CSP was added to perform the competence assays, which, as mentioned in 3.5.4, should mask the effect of Cbf1. Overall, the involvement of *cbf1* in the competence regulon may be relevant for the *in vivo* survival of the pneumococcus: A global transposon-sequencing study revealed that disruption of the *cbf1* gene renders the pneumococcus unable to infect the murine nasopharynx and strongly attenuates murine lung colonization (van Opijnen and Camilli, 2012).

While clearly linked to competence, transcription of the csRNAs is actually upregulated by the CiaRH two-component system (Halfmann et al., 2007), which is not directly involved in the competence regulon. Instead, it is the main regulator protecting the pneumococcus against stress via upregulation of chaperones and heat shock proteins (Gómez-Mejia et al., 2018). The csRNAs and HtrA, a protease that is able to degrade extracellular CSP (Cassone et al., 2012), are among those upregulated genes and together are able to downregulate competence (Gómez-Mejia et al., 2018; Halfmann et al., 2007; Sebert et al., 2005), preventing the activation of the costly competent state under detrimental conditions. It is therefore not surprising that, while constant levels are kept, transcription of the csRNAs is not upregulated under competence-inducing conditions (Slager et al., 2019). Interestingly, this is reminiscent of the quorum sensing system of *Vibrio harveyi*, which is controlled additively by the five Qrr sRNAs (Feng et al., 2015; Pérez-Reytor et al., 2016): even though the csRNAs are highly similar in sequence, they are only partially redundant, as at least
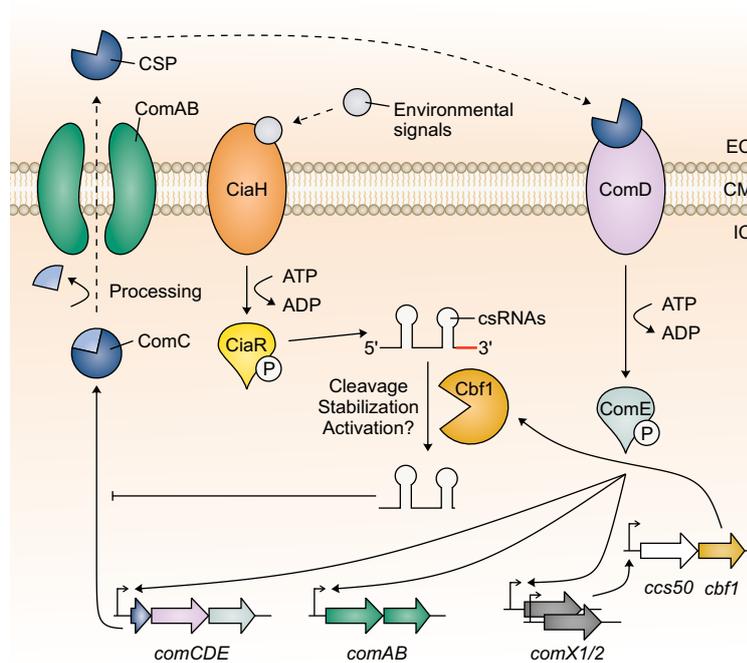
Figure 3.19: The role of Cbf1 in the competence regulon of *S. pneumoniae*. The pneumococcal competence regulon is a quorum sensing system activated by the competence stimulating peptide (CSP). CSP induces the two-component system (TCS) ComDE, which then activates expression of the early competence loci *comAB*, *comCDE*, *comX1* and *comX2*. This results in a positive feedback loop producing ComC, which gets processed to CSP and exported by ComAB. ComX1 and ComX2 are paralogous alternative σ-factors activating the late competence genes, which are, among other things, responsible for DNA uptake and integration. *cbf1* is part of the late competence genes and its protein product is a 3'→5' exonuclease able to trim and thereby stabilize csRNAs, which are regulatory RNAs that inhibit competence by post-transcriptional repression of *comC* translation. Expression of the csRNAs is activated by the TCS CiaRH that is triggered by a variety of environmental signals. EC, extracellular. CM, cell membrane. IC, intracellular.

three of them are necessary to successfully inhibit competence (Laux et al., 2015). To make it even more complicated, it appears that not all combinations of three csRNAs suffice to fulfill this function (Laux et al., 2015).

In summary, the results of this chapter provide a valuable resource for the investigation of RNA and protein complexes in *S. pneumoniae* TIGR4 and other related Gram-positive bacteria. A functional relationship between the exonuclease Cbf1 and the competence-regulating csRNAs was established, resulting in a model in which Cbf1 is upregulated under competence-inducing conditions, leading to stabilization of the available pool of inhibitory csRNAs. On the level of a pneumococcal population, the csRNAs are then able to dampen the energetically costly competence state, possibly by preventing individual cells from entering it in the first place (Figure 3.19). This process might be independent of an RNA chaperone that aids in the facilitation of the pairing between the csRNAs and their target(s).

Chapter 4

Conclusions & Outlook

The original "central dogma" of molecular biology postulated by Francis Crick states that information within the cell can be transferred from nucleic acid to nucleic acid and from nucleic acid to protein but not from protein to protein or from protein to nucleic acid (Crick, 1958, 1970). While this is still regarded as true, another version of the "central dogma", postulated by James Watson, states that information within the cell flows from DNA to RNA and from RNA to protein (Watson, 1965). This second version is the more popular of the two and is still found in many textbooks. However, it is incorrect, as shown by, for example, the discovery of RNA-dependent DNA polymerases (reverse transcriptases) (Temin and Mizutani, 1970). It further generated the general assumption that RNA is only a mere helper of gene expression represented by mRNAs, tRNAs and rRNAs.

This view drastically changed with the discovery of additional classes of ncRNAs. In the 1980s, small nuclear RNAs were discovered as part of the spliceosome and were soon joined by others such as small nucleolar RNAs or microRNAs (Cech and Steitz, 2014). With the introduction of deep sequencing methods, the number of known classes of ncRNAs has dramatically increased and ncRNAs were found to be present in all domains of life (Cech and Steitz, 2014). RNA was further shown to catalyze reactions (called ribozymes) (Kruger et al., 1982), carry all kinds of modifications (Helm and Motorin, 2017) and regulate gene expression (Hör et al., 2020b), just to name a few findings. Importantly, many of these ncRNAs are involved in complexes with proteins to form central RNPs such as the spliceosome (Wilkinson

et al., 2019) or the ribosome (Davis and Williamson, 2017). The formation of RNPs is also important for the function of many if not most of the known post-transcriptional regulators found in bacteria (Babitzke et al., 2019; Holmqvist and Vogel, 2018; Hör et al., 2020b).

Yet, these complexes have been studied on a case-by-case basis, whereas global description of RNA complex formation has not been possible before the introduction of Grad-seq investigating the RNA/protein complexome of *Salmonella* (Smirnov et al., 2016). For the first time, Grad-seq allowed the prediction of complex involvement for all detectable RNAs and proteins based on a single experiment, thereby filling a methodological gap in RNA biology. With the introduction of two additional Grad-seq datasets for *E. coli* MG1655 and *S. pneumoniae* TIGR4 in this thesis, new avenues for research in the field of bacterial RNA biology and complexomics have been opened up.

The availability of Grad-seq datasets for *Salmonella* and *E. coli* now enables the comparison of the complexomes of these closely related enterobacterial species. Importantly, this allows the prediction and verification of conserved functions of RNAs and proteins based on their observed sedimentation profiles. For example, in comparison to Hfq or ProQ, CsrA and its titrating sRNAs CsrB and CsrC do not show any ribosome association in *E. coli* (Figure 2.15) or *Salmonella* (Smirnov et al., 2016). These findings are in line with the function of CsrA to inhibit binding of the ribosome to mRNAs (Romeo and Babitzke, 2018). The combined analysis of the two datasets further multiplies the power of the single datasets: if unexpected or new observations such as the ribosome-association of ProQ (discussed in 2.10) are found in both datasets, the likelihood of them being artifacts is drastically reduced.

Finally, findings in one organism can be cross-referenced with the data of the other organism in order to make functional predictions. For example, the small protein YggL that was identified here to be a ribosome-bound protein in *E. coli* (see 2.9) is also highly conserved in *Salmonella* (Figure 2.26 B). Even though its exact functions remain unknown, *Salmonella* YggL showed the same peak at the 50S subunit observed in *E. coli* (Smirnov et al., 2016). This strongly suggests that YggL has the same functions in both species even though there are no experimental studies on the *Salmonella* homolog.

Similarly, the RNA/protein complexome dataset for *S. pneumoniae* presented in this thesis for the first time allows the global prediction of interactions in a Gram-positive bacterium based on a single experiment. This is of special importance given the lack of studies on Gram-positive RNA and protein biology when compared to Gram-negatives. Grouping of pneumococcal ncRNAs according to their sedimentation behavior followed by tag-based protein capture revealed the 3′→5′ exonuclease Cbf1 to be an interactor of the competence-regulating csRNAs (Figure 3.14). As discussed in 3.6, the proteins recovered in these experiments hint at the absence of a global RBP in the pneumococcus—and possibly other Gram-positives—although it expresses a vast number of regulatory RNAs (Wilton et al., 2015). Importantly, these findings can further be transferred to many related Gram-positive species, which possess homologs of the detected pneumococcal RNAs and proteins. For example, the 6S RNA-RNAP complex was discovered in *E. coli* (Wassarman and Storz, 2000) and described in *B. subtilis* (Burenina et al., 2014; Cavanagh et al., 2012; Trotochaud and Wassarman, 2005). Grad-seq now confirmed its existence in the pneumococcus (Figure 3.6 B), cementing it as a ubiquitous RNP similar to RNase P, SRP, tmRNA-SmpB or the ribosome.

Not only is the knowledge about their complexomes lagging behind, the pneumococcus and other Gram-positives also remain understudied with respect to the functions of their genes. Grad-seq provides a valuable resource for the study of such genes of unknown function and can greatly benefit from the research performed in other organisms: as discussed in 3.6, the sedimentation profiles of SP_0007 and SP_1969 in *S. pneumoniae* matched the functions of their *E. coli* homologs and thereby supported the functional predictions of these proteins. This gives the opportunity to get a quick impression of whether the function of a specific gene of interest might be conserved. If no homologs of an RNA or protein of interest have been functionally characterized, Grad-seq can still provide information about possible interactions and with that add to the results from other experiments such as affinity purification.

To further improve the significance of Grad-seq predictions, it can be combined with synergistic methods. One especially interesting method is R-DeeP (Caudron-Herger et al., 2019, 2020) and the conceptionally similar RNase-sensitive gradient profiles (GradR) (Gerovac, El Mouali, Barquist and Vogel, unpublished). Both meth-

ods rely on the comparison of protein sedimentation profiles between an RNase-treated gradient and a control gradient. Generally, a protein that is either directly or indirectly in contact with RNA will shift toward the top of the gradient upon RNase treatment, thereby enabling the global identification of such proteins. In comparison to the methods discussed in 1.2, R-DeeP/GradR do not identify the RNAs bound to a shifting protein. Grad-seq, however, can provide information on which classes of RNAs might be bound by the shifting proteins, showing how the power of the single methods is increased by their combination.

In the future, the most important step to improve Grad-seq will be to obtain interactome maps with higher resolution. This would allow both to increase the sharpness of sedimentation peaks as well as to better separate similarly sedimenting complexes from each other. Ultimately, an ambitious goal would be to have only few different RNAs and/or proteins per fraction, allowing the direct prediction of interactions between them. In its current iteration, Grad-seq relies on glycerol gradient separation, which does not have the resolution to achieve a fine enough separation for this (see 1.2.5). While an increase in fraction numbers will be able to increase the resolution, detection limits of the downstream high-throughput methods have to be kept in mind.

Therefore, a promising method is to combine several different biochemical separation techniques in order to achieve extremely fine-grained co-elution profiles. This approach was applied to human cells and, using a combination of ion exchange chromatography, isoelectric focusing and sucrose gradients, separated the cellular content into >1,000 biochemical fractions (Havugimana et al., 2012). While this study focused exclusively on protein complexes, it should be possible to use a similar approach to investigate the co-elution of both proteins and RNA as well. Furthermore, an increased resolution would enable the study of more complex samples such as eukaryotic cells or mixtures of bacteria from, *e.g.*, mixed biofilms or the microbiota.

Apart from the obvious increase in experimental costs and time, such a drastic increase in fractions and resolution for Grad-seq will necessitate the development of novel bioinformatical analysis methods to reliably make sense of the data. These pipelines will have to combine the data from the different separation techniques, robustly identify co-eluting molecules and finally visualize them in an easily un-

derstandable fashion. Especially the visualization is key to the success of such a high-resolution Grad-seq variant since it should allow the community to grasp the complexity of the dataset in an intuitive manner. Current complex datasets with hundreds of thousands of data points such as the mouse cell atlas (Han et al., 2018) or the *Tabula Muris* (The Tabula Muris Consortium, 2018), two single cell RNA-seq studies investigating the cellular composition of mouse organs, show examples of how this can be achieved. Yet, to match the requirements for an intuitive visualization of high-resolution Grad-seq, different avenues have to be taken. For example, virtual reality could be an option to allow the three-dimensional exploration of the data, enabling the user to interactively study the similarities of RNA and protein elution profiles.

In addition to increasing the resolution, another important improvement to Grad-seq will be the refinement of the RNA-seq protocols. While the changes introduced in this thesis (see 2.3) improved the ability to normalize the data, the RNA-seq of a Grad-seq experiment is still the step that causes most of the issues. One reason for this is that each fraction of the gradient inevitably has a unique composition of RNAs (see Appendix Figure 7.1). During library preparation, this can cause biases that are different from fraction to fraction during, for example, adapter ligation or PCR amplification (van Dijk et al., 2014). Different options exist to reduce bias during library preparation, *e.g.*, unique molecular identifiers allow the counting of single molecules (Kivioja et al., 2011). Yet, for Grad-seq, empirical studies will be necessary to find the optimal way of reducing biases.

The sequencing platform used for Grad-seq also has to be taken into consideration in order to improve the power of the RNA-seq. The Illumina platform, which has become the industry standard by this point, sequences short reads obtained from cDNA originating from fragmented RNA. While generally of little concern, for Grad-seq this means that discrimination between full-length and processed RNA molecules is not possible based on the sequencing results. However, especially for mRNAs, this knowledge would add an interesting layer to the data: Have mRNA molecules found within the gradient just dissociated from ribosomes? Are they non-functional decay products? Or are they processed but stable, inactive species? Stable 3′ UTR-derived sRNAs (Miyakoshi et al., 2015) already give a flavor of the impor-

tance of this information—they often bind Hfq, which is why they exhibit sedimentation profiles distinct from their usually much less abundant parental mRNAs (Smirnov et al., 2017a). This makes 3' UTR-derived sRNAs easy to detect by Illumina sequencing, which, however, is in contrast to full-length mRNAs.

Third generation sequencing technologies offered by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) provide the ability to sequence full-length single molecules. In comparison to Illumina, they are low-throughput and therefore not suitable for the quantitative analysis needed for Grad-seq (Weirather et al., 2017). Still, PacBio or ONT could be used as complementary methods to add a qualitative layer uncovering the identity of the transcript species within Grad-seq fractions.

Taken together, there are many exciting ways how Grad-seq could develop in the years to come. It will continue to provide a fast and effective way to obtain complexomic resources that should be collated in public databases allowing easy access for everyone in the community.

Chapter 5

Materials & Methods

General chemicals used in this thesis were purchased from AppliChem, Merck, Roth and Sigma.

5.1 General equipment, consumables and software

Table 5.1: General equipment and instruments.

Equipment and instruments	Manufacturer	Model
Biosafety cabinet	Thermo Fisher Scientific	Safe 2020
Cell density meter	Amersham Biosciences	Ultrospec 10
Cell lysis instrument	MP Biomedicals	FastPrep-24
Centrifuge	Eppendorf	5415R and 5424
Centrifuge	Thermo Fisher Scientific	Heraeus Multifuge X3R and Pico 21
Confocal microscope	Leica	TCS SP5
Electrophoresis power supply	Peqlab	EV202, EV232, peqPower 250V/300V
Electroporator	Bio-Rad	MicroPulser
Finescale	Chyo	JL-180
Fraction collector	Gilson	FC 203B
Geiger counter	Thermo Fisher Scientific	Mini 900 Ratemeter
Gel documentation system	Intas	GelStick Imager
Gel dryer	Bio-Rad	Model 583
Gradient profiler	Biocomp	Model 251
Gradient station	Biocomp	Model 153
Horizontal electrophoresis	Peqlab	

Table 5.1, *continued.*

Equipment and instruments	Manufacturer	Model
Hybridization oven	UVP	HB-1000 Hybridizer
Image eraser	GE Healthcare	FLA Image Eraser
Imaging system	GE Healthcare	ImageQuant LAS 4000
Incubator shaker	New Brunswick Scientific	Innova 44
Magnetic stirrer	Wisd Instruments	WiseStir MSH-20A
Mixer mill	Retsch	MM400
Phosphoimager	GE Healthcare	Typhoon FLA 7000
Phosphor screen	Fujifilm	BAS-SR 2040
Pipet aid	Brand	accu-jet pro
Pipettes	Eppendorf	Research plus
Platereader	Tecan	Infinite M Plex
Rotator	Stuart	SB2
RT-qPCR system	Bio-Rad	CFX96
Scale	Kern	572
Sequencing gel chamber	C.B.S. Scientific Co.	SG-400-20
Shaker incubator	Eppendorf	Thermomixer comfort
Sonication system	Diagenode	Bioruptor Plus
Spectrophotometer	Thermo Fisher Scientific	NanoDrop 2000
Thermal cycler	Applied Biosystems	2720
Ultracentrifuge	Beckman Coulter	Optima XP-80
Ultracentrifuge rotor	Beckman Coulter	SW 40 Ti and type 70 Ti
UV crosslinker	Vilber	
UV monitor	Bio-Rad	Econo UV monitor
Vertical electrophoresis	Peqlab	
Vortexer	Scientific Industries	Vortex Genie 2
Waterbath	GFL	1092

Table 5.2: Consumables and commercial kits.

Consumables and kits	Manufacturer
96-well plates	Nunc
Bolt 4–12 % Bis-Tris gels	Thermo Fisher Scientific
Centrifuge tubes	Sarstedt
Cuvettes	Sarstedt
Disposable glass pipettes	Kimble
DNA Cycle Sequencing kit	Jena Bioscience
Electroporation cuvettes	Cell projects

Table 5.2, *continued.*

Consumables and kits	Manufacturer
FastPrep tubes	MP Biomedicals
G-25 MicroSpin columns	GE Healthcare
Glass beads, 0.1 mm	Roth
Hybond+ membranes	GE Healthcare
NucleoSpin Gel and PCR clean-up	Macherey-Nagel
NucleoSpin Plasmid EasyPure	Macherey-Nagel
PCR tubes	Thermo Fisher Scientific
Phase Lock Gel tubes	5 Prime
Pipette tips	Sarstedt
Power SYBR Green RNA-to-CT 1-Step kit	Thermo Fisher Scientific
PVDF membranes	GE Healthcare
StrataClone PCR cloning kit	Agilent
TranscriptAid *in vitro* transcription kit	Thermo Fisher Scientific
Ultracentrifugation tubes	Seton

Table 5.3: Chemicals and commercial reagents.

Chemicals and commercial reagents	Manufacturer
Albumin fraction V	Roth
Ampicillin	Roth
Carbenicillin	Roth
Chloramphenicol	Roth
CSP-1 and CSP-2	AnaSpec
DNA and RNA ladders	Thermo Fisher Scientific
DNA loading buffer ($6\times$)	Thermo Fisher Scientific
dNTPs	Thermo Fisher Scientific
Dynabeads M-270	Thermo Fisher Scientific
Dynabeads protein A/G	Thermo Fisher Scientific
ERCC spike-in	Thermo Fisher Scientific
GlycoBlue	Thermo Fisher Scientific
Kanamycin	Roth
LDS sample buffer	Thermo Fisher Scientific
MES buffer	Thermo Fisher Scientific
Protein ladders	Thermo Fisher Scientific
Rifampicin	Fluka
Roti Hybri-Quick	Roth
Rotiphorese gel 40 (19:1)	Roth

Table 5.3, *continued.*

Chemicals and commercial reagents	Manufacturer
Rotiphorese gel 40 (37.5:1)	Roth
SimplyBlue Coomassie	Thermo Fisher Scientific
TRIzol	Thermo Fisher Scientific
UPS2 spike-in	Sigma
ECL WB detection reagent	GE Healthcare
Yeast extract	Roth

Table 5.4: Enzymes.

Enzymes	Manufacturer
Calf intestinal phosphatase (CIP)	NEB
DNase I	Thermo Fisher Scientific
Lysozyme	Roth
Phusion DNA polymerase	NEB
Polynucleotide kinase (PNK)	Thermo Fisher Scientific
Restriction enzymes	Thermo Fisher Scientific
RNase inhibitor	Thermo Fisher Scientific
Shrimp alkaline phosphatase (SAP)	NEB
SuperScript II reverse transcriptase	Thermo Fisher Scientific
T4 DNA ligase	NEB
Taq DNA polymerase	NEB

Table 5.5: Antibodies and antisera.

Antibodies and antisera	Dilution for WB	Source
α-6xHis (mouse)	1:3,000	Sigma
α-FLAG (mouse)	1:1,000	Sigma
α-GAPDH (rabbit)	1:2,000	B. Henriques-Normark (Stockholm)
α-GroEL (rabbit)	1:10,000	Sigma
α-mouse; HRP-conjugated (goat)	1:10,000	Thermo Fisher Scientific
α-rabbit; HRP-conjugated (goat)	1:10,000	Thermo Fisher Scientific
α-RpoB (mouse)	1:1,000	BioLegend
α-RpoD (mouse)	1:1,000	BioLegend

Table 5.6: Software.

Software and version	Purpose	Developer/Reference
CorelDRAW 2018	Vector graphics editor	Corel Corporation
Excel 365 ProPlus	Spreadsheet editor	Microsoft Corporation
ImageJ 1.52u	Image processing	Schneider et al. (2012)
IGV 2.8.2	Genome browser	Thorvaldsdóttir et al. (2013)
JabRef 4.3.1	Reference manager	The JabRef Team
Jalview 2.11	Multiple sequence alignment visualization	Waterhouse et al. (2009)
MiKTeX 2.9.7300	LaTeX distribution	Christian Schenk
Notepad++ 7.7.1	Text editor	Notepad++ team
Prism 8.4.0	Graphing and statistics	GraphPad Software, Inc.
Texmaker 5.0.4	LaTeX editor	Pascal Brachet
VARNA 3-93	RNA secondary structure visualization	Darty et al. (2009)

5.2 Solutions and media

Table 5.7: Buffers and solutions.

Buffers and solutions	Composition
30:1 mix	29 ml ethanol, 1 ml 3 M sodium acetate, pH 5.2 or 6.5
DEPC-H_2O	1 ml DEPC in 1 l H_2O; 37 °C o/n; autoclave
Developer solution	60 g Na_2CO_3; 4 mg $Na_2S_2O_3 \times 5\,H_2O$; 0.5 ml formaldehyde (37 %); H_2O ad 1 l
Fixing solution	500 ml ethanol; 120 ml acetic acid; 0.5 ml formaldehyde (37 %); H_2O ad 1 l
Glycerol solution (10 %)	10 g glycerol in lysis buffer A + 1 mM PMSF and 0.2 % (v/v) Triton X 100
Glycerol solution (40 %)	40 g glycerol in lysis buffer A + 1 mM PMSF and 0.2 % (v/v) Triton X 100
Lower buffer (SDS-PAGE)	1.5 M Tris-HCl, pH 8.8, 0.4 % (w/v) SDS
Lysis buffer A	20 mM Tris-HCl, pH 7.5; 150 mM KCl; 1 mM $MgCl_2$; 1 mM DTT
Lysis buffer B	20 mM Tris-HCl, pH 7.5; 100 mM NH_4Cl; 10.5 mM $MgCl_2$; 0.5 mM EDTA; 3 mM DTT
Lysis buffer C	20 mM Tris-HCl, pH 7.5; 100 mM NH_4Cl; 10 mM $MgCl_2$; 1 mM DTT
Lysis buffer D	20 mM Tris-HCl, pH 7.5; 100 mM NH_4Cl; 10 mM $MgCl_2$; 3 mM DTT
Lysis buffer E	50 mM Tris-HCl, pH 8; 150 mM KCl; 1 mM $MgCl_2$; 5 % (v/v) glycerol; 0.05 % (v/v) Tween-20
Lysis buffer F	50 mM Tris-HCl, pH 8; 150 mM KCl; 1 mM $MgCl_2$; 5 % (v/v) glycerol; 0.05 % (v/v) Tween-20; 1 mM PMSF; 1 mM DTT
MOPS buffer (10×)	41.8 g MOPS; 16.6 ml 3 M sodium acetate; 20 ml 0.5 M EDTA, pH 8; H_2O ad 1 l; adjust pH to 7 with NaOH

Table 5.7, *continued.*

Buffers and solutions	Composition
P1 buffer	10 mM MgSO$_4$; 5 mM CaCl$_2$
PAA solution (denaturing RNA gel)	100 ml 10× TBE; 420 g urea; 150 or 250 ml Rotiphorese gel 40 (19:1) for 6 or 10 % gels; H$_2$O ad 1 l
PAA solution (protein resolving gel)	3.75 ml lower buffer; 3 or 3.25 ml Rotiphorese gel 40 (37.5:1) and 3.25 or 3 ml H$_2$O for 12 or 15 % gels; 75 µl 10 % (w/v) APS; 7.5 µl TEMED
PAA solution (protein stacking gel)	1.25 ml upper buffer; 1 ml Rotiphorese gel 40 (37.5:1); 7.5 ml H$_2$O; 90 µl 10 % (w/v) APS; 9 µl TEMED
PBS (10×)	2 g KCl; 2.4 g KH$_2$PO$_4$; 80 g NaCl; 14.4 g Na$_2$HPO$_4$; adjust to pH 7.4; H$_2$O ad 1 l
Protein loading buffer (5×)	15 g SDS; 46.95 ml 1 M Tris-HCl, pH 6.8; 75 ml glycerol; 11.56 g DTT; 0.075 g bromophenol blue; H$_2$O ad 150 ml
RNA elution buffer	0.1 M sodium acetate, pH 6.5; 0.1 % (w/v) SDS; 10 mM EDTA, pH 8
RNA loading buffer (2×)	0.025 % (w/v) bromophenol blue; 0.025 % (w/v) xylene cyanol; 18 µM EDTA, pH 8; 0.13 % (w/v) SDS; 95 % foramamide
SB 1× Mg10	10 mM Tris-acetate, pH 7.6; 100 mM potassium acetate; 1 mM DTT; 10 mM magnesium acetate
SB 1× Mg60	10 mM Tris-acetate, pH 7.6; 100 mM potassium acetate; 1 mM DTT; 60 mM magnesium acetate
SB 5× –Mg	50 mM Tris-acetate, pH7.6; 500 mM potassium acetate; 5 mM DTT
SDS running buffer (10×)	30.275 g Tris base; 144 g glycine; 10 g SDS; H$_2$O ad 1 l
Sensitizer	0.2 g Na$_2$S$_2$O$_3$ × 5 H$_2$O; H$_2$O ad 1 l
Silver staining solution	2 g AgNO$_3$; 0.75 ml formaldehyde (37 %); H$_2$O ad 1 l
Silver stop solution	10 g glycine; H$_2$O ad 1 l
SSC (20×)	3 M NaCl; 0.3 M sodium citrate, pH 7
SSC-S	Needed dilution of SSC + 0.1 % (w/v) SDS
Storage buffer	Lysis buffer B + 10 % (v/v) glycerol
Stop mix	95 % ethanol, 5 % acidic phenol
TAE (50×)	242 g Tris base; 51.7 ml acetic acid; 10 mM EDTA, pH 8; H$_2$O ad 1 l
TBE (10×)	108 g Tris base; 55 g boric acid; 20 mM EDTA, pH 8; H$_2$O ad 1 l
TBS (10×)	24.11 g Tris base; 72.6 g NaCl; adjust to pH 7.4 with HCl; H$_2$O ad 1 l
TBS-T	1× TBS + 0.1 % (v/v) Tween-20
TE (1×)	100 mM Tris-HCl, pH 8; 10 mM EDTA, pH 8
Toeprint stop buffer	50 mM Tris-HCl, pH 7.5; 0.1 % (w/v) SDS; 10 mM EDTA, pH 8
Transfer buffer	3 g Tris base; 14.4 g glycine; 200 ml methanol; H$_2$O ad 1 l
Upper buffer (SDS-PAGE)	0.5 M Tris-HCl, pH 6.8, 0.4 % (w/v) SDS
Wash buffer A	Lysis buffer F with a total of 300 mM KCl
Wash buffer B	Lysis buffer F + 0.1 % (v/v) Triton X 100

Table 5.8: Media.

Media	Composition
Lennox broth (LB)	10 g tryptone; 5 g yeast extract; 5 g NaCl; H_2O ad 1 l
LB agar	10 g tryptone; 5 g yeast extract; 5 g NaCl; 15 g agar; H_2O ad 1 l
Todd Hewitt broth + yeast extract (THY)	30 g Todd Hewitt broth (Bacto); 5 g yeast extract; H_2O ad 1 l
Tryptic soy agar with sheep blood (TSA-B)	15 g tryptone; 5 g soy peptone; 5 g NaCl; 15 g agar; 50 ml defibrinated sheep blood; H_2O ad 1 l (bought from Oxoid)

5.3 Bacterial strains, plasmids and oligonucleotides

Table 5.9: Bacterial strains.

Name	Species	Genotype	Plasmid	Source
JVS-00271	*S.* Typhimurium SL1344	wild type	pJV300	Laboratory strain collection
JVS-01574	*S.* Typhimurium SL1344	wild type		Laboratory strain collection
JVS-04335	*E. coli* BL21(DE3)	wild type		Laboratory strain collection
JVS-05709	*E. coli* MG1655	wild type		Laboratory strain collection
JVS-10520	*E. coli* W3110	*proQ*-3xFLAG-KmR		Holmqvist et al. (2018)
JVS-11947	*E. coli* MG1655	wild type	pKD46	This study
JVS-11972	*E. coli* MG1655	*intS*-3xFLAG-KmR		This study
JVS-11975	*E. coli* BL21(DE3)	wild type	pJV300	This study
JVS-11986	*E. coli* MG1655	wild type	pJV300	This study
JVS-12009	*S. pneumoniae* TIGR4	wild type		B. Henriques-Normark
JVS-12010	*S. pneumoniae* TIGR4	Δ*cbf1*::Sp		B. Henriques-Normark
JVS-12012	*S. pneumoniae* TIGR4	*cbf1*-3xFLAG-Sp		B. Henriques-Normark
JVS-12013	*S. pneumoniae* TIGR4	Δ*cbf1*::Sp Δ*bgaA*::[*tetM*-P$_{czcD}$-*cbf1*]		B. Henriques-Normark
JVS-12054	*E. coli* 536	wild type		Laboratory strain collection
JVS-12065	*E. coli* 536	wild type	pJV300	This study
JVS-12077	*E. coli* MG1655	Δ*yggL*	pJV300	This study
JVS-12078	*E. coli* MG1655	Δ*yggL*	pJH034	This study
JVS-12082	*E. coli* MG1655	Δ*ryeG*		This study
JVS-12083	*E. coli* MG1655	Δ*ryeG*	pJV300	This study
JVS-12084	*E. coli* MG1655	Δ*ryeG*	pJH032	This study
JVS-12107	*S.* Typhimurium SL1344	wild type	pJH032	This study
JVS-12108	*E. coli* BL21(DE3)	wild type	pJH032	This study

Table 5.9, *continued.*

Name	Species	Genotype	Plasmid	Source
JVS-12109	*E. coli* 536	wild type	pJH032	This study
JVS-12151	*E. coli* MG1655	*yggL* truncation		This study
JVS-12156	*E. coli* MG1655	Δ*yggL*		This study
JVS-12157	*E. coli* MG1655	*yggL*-3xFLAG		This study
JVS-12223	*E. coli* MG1655	wild type	pJH066	This study
JVS-12224	*E. coli* MG1655	*yggL* truncation	pJH066	This study
JVS-12225	*E. coli* MG1655	Δ*yggL*	pJH066	This study
JVS-12226	*E. coli* MG1655	*yggL* truncation	pJH067	This study
JVS-12227	*E. coli* MG1655	Δ*yggL*	pJH067	This study
JVS-12246	*E. coli* MG1655	Δ*ryeG*	pJH071	This study
JVS-12262	*E. coli* MG1655	Δ*ryeG*	pJH068	This study
JVS-12277	*E. coli* MG1655	Δ*ryeG*	pJH078	This study
JVS-12278	*E. coli* MG1655	Δ*ryeG*	pJH076	This study
JVS-12279	*E. coli* MG1655	Δ*ryeG*	pJH075	This study
JVS-12419	*S. pneumoniae* R6	wild type		B. Henriques-Normark
JVS-12420	*S. pneumoniae* R6	Δ*cbf1*::Sp		B. Henriques-Normark

Table 5.10: Plasmids.

Name	Resistance	Backbone	Purpose	Source
pCP20	Amp; Cm		expresses Flp recombinase	Cherepanov and Wackernagel (1995)
pJH032	Amp	pZE12	RyeG overexpression	this study
pJH034	Amp	pZE12	*yggL* overexpression	this study
pJH066	Cm	pXG-1	ctrl plasmid	this study
pJH067	Cm	pXG-1	*yggL* expression (own promoter)	this study
pJH068	Amp	pCRT7	RyeG inducible overexpression	this study
pJH071	Amp	pCRT7	ctrl plasmid	this study
pJH075	Amp	pCRT7	RyeG SD-mut inducible overexpression	this study
pJH076	Amp	pCRT7	RyeG ORF3-stop inducible overexpression	this study
pJH078	Amp	pCRT7	RyeG-ORF2-stop inducible overexpression	this study
pJV300	Amp	pZE12	ctrl plasmid	Sittka et al. (2007)
pKD4	Amp		template for Km knockouts	Datsenko and Wanner (2000)
pKD46	Amp		expresses λRED recombinase	Datsenko and Wanner (2000)

Table 5.10, *continued.*

Name	Resistance	Backbone	Purpose	Source
pSUB11	Amp		template for 3xFLAG-tagging	Uzzau et al. (2001)

Table 5.11: Oligonucleotides.

Name	Sequence (5'→3')	Purpose
JVO-00249	GGTATTAGCTACCGTTTCCAG	NB probe for E. coli 16S rRNA
JVO-00321	CAATTGCAAACACAACAACACAACATC	NB probe for E.coli GcvB
JVO-00322	CTACGGCGTTTCACTTCTGAGTTC	NB probe for E. coli 5S rRNA
JVO-01003	GAATCTCCGAGATGCCG	NB probe for E. coli 6S RNA
JVO-01341	AATACGGCGCAGTGCGTTA	NB probe for E. coli rpoA
JVO-01366	CACCAATACTCAGTCACACATGATG	NB probe for E. coli SgrS
JVO-01367	TCGTTACACCAGGAAATCTGATGTG	NB probe for E. coli DsrA
JVO-01368	GGTCTGAAAGATAGAACATCTTACCTCTGTACC	NB probe for E. coli Spot 42
JVO-01370	GTTGTTTCACTCAGGGGATTTCCATG	NB probe for E. coli RprA
JVO-01371	GGATGATGATAACAAATGCGCGTCT	NB probe for E. coli MicA
JVO-01727	CTATTGGCCCGTCAAAGAGGAATTTCA	NB probe for E. coli ChiX
JVO-02280	TGGCACCGTTAGTGATTTCGA	NB probe for E. coli cspE
JVO-02873	GGTTCCTGGTACAGCTAGCA	NB probe for E. coli CyaR
JVO-04201	GTTTTTTTTAATACGACTCACTATAGGTCGTACACCATCAGGGTAC	Sense oligo for MS2-tag T7 template
JVO-04202	GTTTTTTTTAATACGACTCACTATAGG	Sense oligo for T7 promotor
JVO-04203	GTGACCAGACCCTGATGG	Antisense oligo for MS2 with linker
JVO-04943	ACAGACCCTGATGGTGTCT	Antisense oligo for MS2 with U (for 3'tagging)
JVO-04944	GTCACCGTACACCATCAGGGTAC	Sense oligo for MS2 with linker (for 3'tagging)
JVO-08219	GGCCCTGCCAGCTACAT	NB probe for E. coli 4.5S RNA
JVO-08540	CCTCCGACCCCTTCG	NB probe for E. coli tRNA(Pro)CGG
JVO-09045	AACGGTCTGACCTTCGCTTACTCG	NB probe for E. coli rplU
JVO-10325	TGGTGGAGCTGGCGGGAGTT	NB probe for E. coli tmRNA
JVO-10329	AGGTGAAACTGACCGATAAGCC	NB probe for E. coli RnpB / M1 RNA
JVO-14661	ACAGCCCCTCGACACATAAC	NB probe for S. pneumoniae tmRNA

Table 5.11, *continued.*

Name	Sequence (5'→3')	Purpose
JVO-14662	TACGTTCGTTTCCGCGTACT	NB probe for S. pneumoniae RNase P
JVO-14665	GGAGTTCATTAAATAAAGATATTAGATGAAAATCAAATTCAAACTAATTC	NB probe for S. pneumoniae csRNA5
JVO-14666	CATGGGTACAGGTGTATCTC	NB probe for S. pneumoniae 5S rRNA
JVO-14667	GGCTACCTCTATCAAGGTGTACTCCTTCTATACTATCCCTTGTGCTTTAG	NB probe for S. pneumoniae F41
JVO-14670	CATTTCACCGCTACACATGG	NB probe for S. pneumoniae 16S rRNA
JVO-14671	CCTGATATACTTCCCTTGGGCTACTAGTCTTTCAGATTCCTATTCAATTAC	NB probe for S. pneumoniae F20
JVO-14673	CGACCTCACGCTTATCAGGC	NB probe for S. pneumoniae tRNA-Ile
JVO-14856	TGAAGCTGTTTCCGTGTGAG	NB probe for S. pneumoniae 6S RNA
JVO-14874	CACCATCAGGGTCTGGTCACGGTTACAAGAAGACCTCCTAACTTGTTG	5' MS2 tagging of S. pneumoniae csRNA3; fwd
JVO-14875	AAAAAAGCCACCGAATGCGGTGAC	5' MS2 tagging of S. pneumoniae csRNA3; rev
JVO-14876	GTTTTTTTTAATACGACTCACTATAGGGGTTACAAGAAGACCTCCTAACTTGTTG	3' MS2 tagging of S. pneumoniae csRNA3; fwd
JVO-14877	GTACCCTGATGGTGTACGGTGACAAAAAAGCCACCGAATGCGGTGAC	3' MS2 tagging of S. pneumoniae csRNA3; rev
JVO-14878	CACCATCAGGGTCTGGTCACATTAAATAAAGACCTCCTAATATTATTTG	5' MS2 tagging of S. pneumoniae csRNA5; fwd
JVO-14879	AATAAAAGCCACCCATACAGGCGAC	5' MS2 tagging of S. pneumoniae csRNA5; rev
JVO-14880	GTTTTTTTTAATACGACTCACTATAGGATTAAATAAAGACCTCCTAATATTATTTG	3' MS2 tagging of S. pneumoniae csRNA5; fwd
JVO-14881	GTACCCTGATGGTGTACGGTGACAATAAAAGCCACCCATACAGGCGAC	3' MS2 tagging of S. pneumoniae csRNA5; rev
JVO-14882	CACCATCAGGGTCTGGTCACATAGAAACGCTGTGGTGTACGACTTC	5' MS2 tagging of S. pneumoniae 6S RNA; fwd
JVO-14883	AAAAAAGCTGTATTGGTGCCGAAAC	5' MS2 tagging of S. pneumoniae 6S RNA; rev
JVO-14884	GTTTTTTTTAATACGACTCACTATAGGATAGAAACGCTGTGGTGTACGACTTC	3' MS2 tagging of S. pneumoniae 6S RNA ; fwd
JVO-14885	GTACCCTGATGGTGTACGGTGACAAAAAAGCTGTATTGGTGCCGAAAC	3' MS2 tagging of S. pneumoniae 6S RNA ; rev
JVO-14886	CACCATCAGGGTCTGGTCACAAAGCACAAGGGATAGTATAGAAGGAG	5' MS2 tagging of S. pneumoniae F41; fwd
JVO-14887	AAAAAGCACCTAGTTTCCTAGATGCTAGC	5' MS2 tagging of S. pneumoniae F41; rev
JVO-14888	GTTTTTTTTAATACGACTCACTATAGGAAAGCACAAGGGATAGTATAGAAGGAG	3' MS2 tagging of S. pneumoniae F41; fwd

Table 5.11, *continued.*

Name	Sequence (5'→3')	Purpose
JVO-14889	GTACCCTGATGGTGTACGGTGACAAAAAGCACCTAGTTTCCTAGATGCTAGC	3' MS2 tagging of S. pneumoniae F41; rev
JVO-14987	CACCATCAGGGTCTGGTCACAATTGAATAGGAATCTGAAAGACTAGTAGC	5' MS2 tagging of S. pneumoniae F20; fwd
JVO-14988	AAAAAGAAAACCACATGCCAAAAACTCCAC	5' MS2 tagging of S. pneumoniae F20; rev
JVO-14989	GTTTTTTTTAATACGACTCACTATAGGAATTGAATAGGAATCTGAAAGACTAGTAGC	3' MS2 tagging of S. pneumoniae F20; fwd
JVO-14990	GTACCCTGATGGTGTACGGTGACAAAAAGAAAACCACATGCCAAAAACTCCAC	3' MS2 tagging of S. pneumoniae F20; rev
JVO-14991	CACCATCAGGGTCTGGTCACGTTAAGTGACGATAGCCTAGGAGATACAC	5' MS2 tagging of S. pneumoniae 5S rRNA; fwd
JVO-14992	GCTAAGCGACTTCCCTATCTCAC	5' MS2 tagging of S. pneumoniae 5S rRNA; rev
JVO-14993	GTTTTTTTTAATACGACTCACTATAGGGTTAAGTGACGATAGCCTAGGAGATACAC	3' MS2 tagging of S. pneumoniae 5S rRNA; fwd
JVO-14994	GTACCCTGATGGTGTACGGTGACGCTAAGCGACTTCCCTATCTCAC	3' MS2 tagging of S. pneumoniae 5S rRNA; rev
JVO-14995	CACCATCAGGGTCTGGTCACGGGGTCGTTACGGATTCGACAGGCATTATG	5' MS2 tagging of S. pneumoniae tmRNA; fwd
JVO-14996	AATGGAGCCGGTGGGAGTCGAACCCACG	5' MS2 tagging of S. pneumoniae tmRNA; rev
JVO-14997	GTTTTTTTTAATACGACTCACTATAGGGGGGTCGTTACGGATTCGACAGGCATTATG	3' MS2 tagging of S. pneumoniae tmRNA; fwd
JVO-14998	GTACCCTGATGGTGTACGGTGACAATGGAGCCGGTGGGAGTCGAACCCACG	3' MS2 tagging of S. pneumoniae tmRNA; rev
JVO-15000	AAGTTTAGGATATTTGTTACAACAAGTTAGGAGGTCTTCTTGTAACC	NB probe for S. pneumoniae csRNA3
JVO-15001	ATTATGAAAAAGTTTTAGGAGTTTAAGTTAAGGTCTTCTTAACTTAT	NB probe for S. pneumoniae csRNA4
JVO-15118	CGTACCCTGATGGTGTACGA	NB Probe against MS2 (first hairpin)
JVO-15295	CACCATCAGGGTCTGGTCACATTAAATAAAGACCTCCTAACTTTATTTA	5' MS2 tagging of S. pneumoniae csRNA1/2; fwd
JVO-15296	AAAAAAGCCACCTGATTGGGTGGCTTCATTAG	5' MS2 tagging of S. pneumoniae csRNA1; rev
JVO-15297	GTTTTTTTTAATACGACTCACTATAGGATTAAATAAAGACCTCCTAACTTTATTTA	3' MS2 tagging of S. pneumoniae csRNA1/2; fwd
JVO-15298	GTACCCTGATGGTGTACGGTGACAAAAAAGCCACCTGATTGGGTGGCTTCATTAG	3' MS2 tagging of S. pneumoniae csRNA1; rev

Table 5.11, *continued.*

Name	Sequence (5'→3')	Purpose
JVO-15299	`AAAAAACTCCACCTGATTCGGTGGAGTTAAG`	5' MS2 tagging of S. pneumoniae csRNA2; rev
JVO-15300	`GTACCCTGATGGTGTACGGTGACAAAAAACTCCACCTGATTCGGTGGAGTTAAG`	3' MS2 tagging of S. pneumoniae csRNA2; rev
JVO-15301	`CACCATCAGGGTCTGGTCACATAAGTTAAGAAGACCTTAACTTAAACTC`	5' MS2 tagging of S. pneumoniae csRNA4; fwd
JVO-15302	`AAAAAATAAGCCGCCTGATTGGGCGAC`	5' MS2 tagging of S. pneumoniae csRNA4; rev
JVO-15303	`GTTTTTTTTAATACGACTCACTATAGGATAAGTTAAGAAGACCTTAACTTAAACTC`	3' MS2 tagging of S. pneumoniae csRNA4; fwd
JVO-15304	`GTACCCTGATGGTGTACGGTGACAAAAAATAAGCCGCCTGATTGGGCGAC`	3' MS2 tagging of S. pneumoniae csRNA4; rev
JVO-15428	`GTTGATTCCATTTCCGTTTAATTAC`	NB probe E. coli CsrC
JVO-15429	`CCTCTGGCCTTGCGGCCAATCGTTC`	NB probe E. coli CsrB
JVO-15448	`GTTTTTTTTAATACGACTCACTATAGGGAGACCTAGCCTATTAAATAAAGACCTCCTAACTTTATTTA`	5' tagging of S. pneumoniae csRNA1/2; fwd; use with JVO-15296/15299
JVO-15449	`GTTTTTTTTAATACGACTCACTATAGGGAGACCTAGCCTGGTTACAAGAAGACCTCCTAACTTGTTG`	5' tagging of S. pneumoniae csRNA3; fwd; use with JVO-14875
JVO-15450	`GTTTTTTTTAATACGACTCACTATAGGGAGACCTAGCCTATAAGTTAAGAAGACCTTAACTTAAACTC`	5' tagging of S. pneumoniae csRNA4; fwd; use with JVO-15302
JVO-15451	`GTTTTTTTTAATACGACTCACTATAGGGAGACCTAGCCTATTAAATAAAGACCTCCTAATATTATTTG`	5' tagging of S. pneumoniae csRNA5; fwd; use with JVO-14879
JVO-15452	`GTTTTTTTTAATACGACTCACTATAGGGAGACCTAGCCTAAAGCACAAGGGATAGTATAGAAGGAG`	5' tagging of S. pneumoniae F41; fwd; use with JVO-14887
JVO-15453	`GTTTTTTTTAATACGACTCACTATAGGGAGACCTAGCCTAATTGAATAGGAATCTGAAAGACTAGTAGC`	5' tagging of S. pneumoniae F20; fwd; use with JVO-14988
JVO-15454	`GTTTTTTTTAATACGACTCACTATAGGGAGACCTAGCCTGGGGTCGTTACGGATTCGACAGGCATTATG`	5' tagging of S. pneumoniae tmRNA; fwd; use with JVO-14996
JVO-15455	`GTTTTTTTTAATACGACTCACTATAGGGAGACCTAGCCTGTTAAGTGACGATAGCCTAGGAGATACAC`	5' tagging of S. pneumoniae 5S rRNA; fwd; use with JVO-14992

Table 5.11, *continued.*

Name	Sequence (5′→3′)	Purpose
JVO-15456	GTTTTTTTTAATACGACTCACTATAGGGAGACCTAGCCTATAGAAACGCTGTG GTGTACGACTTC	5′ tagging of S. pneumoniae 6S RNA; fwd; use with JVO-14883
JVO-16298	TCACGGAGTATGGGACTTAATGTTCGAGAGGGTGAAATAAGTGTAGGCTGGAG CTGCTTC	fwd primer to delete yggL in E. coli using pKD4
JVO-16299	CATCAGGCAGTTTTGCATTTGTCATCGCCCGTATGCTTTCGGTCCATATGAAT ATCCTCCTTAG	rev primer to delete yggL in E. coli using pKD4
JVO-16300	TGAGGTACGCACCAGCGAACTTTTCGACGTTTGGTGGGACGACTACAAAGACC ATGACG	fwd primer to 3xFLAG-tag E. coli yggL
JVO-16301	CATCAGGCAGTTTTGCATTTGTCATCGCCCGTATGCTTTCCATATGAATATCC TCCTTAG	rev primer to 3xFLAG-tag E. coli yggL
JVO-16302	CAATGATTACGTACCGCGTCC	fwd primer to verify yggL deletion
JVO-16303	GCGCATCATAGTCTTCCCTC	rev primer to verify yggL deletion/tagging
JVO-16339	ACTATCAAATGCCCTGATTC	NB E. coli RyeG
JVO-16340	CTGGAATTCGTCGATGTGCA	NB E. coli yggL CDS
JVO-16349	GTTATCTGGCCTGGGAAGGT	fwd primer to verify yggL tagging; use with JVO-16303
JVO-16365	GAGTTGCGAACGGGTGAGTA	fwd 16S rRNA for S. pneumoniae DNase treatment control PCR
JVO-16366	TCAGACTTCCGTCCATTGCC	rev 16S rRNA for S. pneumoniae DNase treatment control PCR
JVO-16367	TCCTGTCTGGCAACGGATTG	fwd cbf1 qPCR S. pneumoniae
JVO-16368	CCGTTTCAAAGGCATGGTGG	rev cbf1 qPCR S. pneumoniae
JVO-16369	CCGTATGGCTCAATGGTGGA	fwd gyrA qPCR S. pneumoniae
JVO-16370	TATAACGTTGAGCGGCAGCA	rev gyrA qPCR S. pneumoniae
JVO-16384	GATGAGACAAAACTTATACACACAAAGC	rev primer to verify intS deletion

Table 5.11, *continued.*

Name	Sequence (5′→3′)	Purpose
JVO-16385	TAGTATATGTATCTATCACTGTTGATGATAATATCAGCACGTGTAGGCTGGAGCTGCTTC	fwd primer to delete ryeG in E. coli using pKD4
JVO-16386	AAATCAAGACAATAGCGATGTAAAGCACTAAGTTTAAATTGGTCCATATGAATATCCTCCTTAG	rev primer to delete ryeG in E. coli using pKD4
JVO-16387	CAAAGTGGAAGTCCTTATGTTAAGTATAA	fwd primer to verify ryeG deletion
JVO-16388	GACCTTTGCTTATACATAACAGTCGT	rev primer to verify ryeG deletion
JVO-16389	GATGCAATGGTGGGCGGACTGGCTTGATGAGAAGGTGGAGGACTACAAAGACCATGACG	fwd primer to 3xFLAG-tag E. coli intS
JVO-16390	ATTGCAAGACTTTGTGCTATTCGATAGTTGTTAAGGTCGCCATATGAATATCCTCCTTAG	rev primer to 3xFLAG-tag E. coli intS
JVO-16391	GGCCTGCTGACGCTATTGAA	fwd primer to verify intS tagging; use with JVO-16384
JVO-16392	P-TTGGTTCTGGAGGGGGTTTGTT	fwd primer to clone RyeG into pZE12; phosphorylated
JVO-16393	GTTTTTCTAGAAAATTGCGCGCCAATCATGGC	rev primer to clone RyeG into pZE12; XbaI
JVO-16396	GTTTTGGTACCATGGCAAAGAACCGTAGCCGT	fwd primer to clone E. coli yggL into pZE12; KpnI
JVO-16397	GTTTTTCTAGATTAGTCCCACCAAACGTCGAAAAG	rev primer to clone E. coli yggL into pZE12; XbaI
JVO-16411	AAATTGCGCGCCAATCATGGC	rev primer for 14mer pulldown of E. coli RyeG
JVO-16446	GTTTTTCTAGACTTATACATAACAGTCGTTTTTTTTAATTTATAAAATAATTC	rev primer to clone RyeG into pZE12 (binds 40nt dwnstr); XbaI; use with JVO-16392
JVO-16459	GTTTTTCTAGACAAAGAAATCAAGACAATAGCGATGTAAA	rev primer to clone RyeG into pZE12 (binds 11nt dwnstr); XbaI; use with JVO-16392
JVO-16507	GTTTTTTTTAATACGACTCACTATAGGGTTGGTTCTGGAGGGGGTTTGT	fwd primer to amplify E. coli RyeG with a T7 promoter; use with JVO-16411

Table 5.11, *continued.*

Name	Sequence (5'→3')	Purpose
JVO-16525	GTTTTGACGTCGCGTGATGTGTCACGATGCG	fwd primer to clone the transcriptional unit of E. coli yggL into pXG-1; AatII
JVO-16526	GTTTTTCTAGAGCGTGACGCTGCACTGGTA	rev primer to clone the transcriptional unit of E. coli yggL into pXG-1; XbaI
JVO-16531	GCTCTTCCAGCCACTTACGCACAATCGCCTGATGTTCTTCGGTCCATATGAAT ATCCTCCTTAG	rev primer to truncate E. coli yggL; keeps transcriptional unit of yggN intact. Use with JVO-16298
JVO-16577	P-CTAGAGGCATCAAATAAAACGAAAGG	fwd primer to generate a ctrl vector of pXG-1; phosphorylated
JVO-16578	GTGCTCAGTATCTCTATCACTGAT	rev primer to generate a ctrl vector of pXG-1; binds the -1 position of the PLtetO-1 promoter
JVO-16602	ATACGCCCGGTAGTGATCTTATT	fwd primer for colony PCR and sequencing in pXG-1 backbone
JVO-16648	GTTTTCTCGAGGCGTGATGTGTCACGATGCG	fwd primer to clone the transcriptional unit of E. coli yggL into pXG-1; XhoI; use with JVO-16526
JVO-16771	P-ATGCATTTAAGTTATCGTCTGCAGATAG	fwd primer to delete SD sequence of RyeG on plasmids; phosphorylated
JVO-16772	CATTGCCCACAACAAACTTTTCCCAGAACCAA	rev primer to delete SD sequence of RyeG on plasmids
JVO-16791	P-GCGTGATGTGTCACGATGCG	fwd primer to clone the transcriptional unit of E. coli yggL into pXG-1; phosphorylated; use with JVO-16526
JVO-16792	CTCGAGGTGAAGACGAAAGG	rev primer to PCR pXG-1 in order to blunt clone inserts; binds at the XhoI site

Table 5.11, *continued.*

Name	Sequence (5′→3′)	Purpose
JVO-16793	GTTTTTCTAGAGGCATCAAATAAAACGAAAGG	fwd primer to PCR pXG-1 in order to blunt clone inserts; binds at the XbaI site and is cleavable
JVO-16794	GTTTTAAGCTTCTAGCATAACCCCTTGGGG	fwd primer to PCR pCRT7; use with JVO-16578 for sRNA cloning
JVO-16795	P-AGCTTCTAGCATAACCCCTTGG	fwd primer to make a ctrl plasmid of pCRT7; phosphorylated; use with JVO-16578
JVO-16796	GTTTTAAGCTTAAATTGCGCGCCAATCATGGC	rev primer to clone RyeG into pCRT7; HindIII; use with JVO-16392
JVO-16797	GTTTTAAGCTTCTTATACATAACAGTCGTTTTTTTAATTTATAAAATAATTC	rev primer to clone RyeG into pCRT7 (binds 40nt downstream of 3′ end); HindIII; use with JVO-16392
JVO-16798	GTTTTAAGCTTCAAAGAAATCAAGACAATAGCGATGTAAA	rev primer to clone RyeG into pCRT7 (binds 11nt downstream of 3′ end); HindIII; use with JVO-16392
JVO-16825	CAGCTCTAATGCGCTGTTAATCAC	fwd primer for colony PCR and sequencing of pCRT7
JVO-16826	ATCGGTGCGGGCCTCTTC	rev primer for colony PCR and sequencing of pCRT7
JVO-16833	ACCAACCTAATTTCTCCTCCTC	rev primer for toe printing of RyeG; binds 60 nt from the 3′ end
JVO-16871	P-CAATGATGCATTAAAGTTATCGTCTGC	fwd primer to introduce stop codon into RyeG ORF3 on plasmids; phosphorylated; use with JVO-16550
JVO-16872	GTTTTGGTACCTTGTGGGCAATGATGCA	fwd primer to clone RyeG ORF3 into pZE12/pCRT7; KpnI
JVO-16873	GTTTTTCTAGATTATTGTAATATCTCCTCTATC	rev primer to clone RyeG ORF3 into pZE12; XbaI
JVO-16874	GTTTTAAGCTTTTATTGTAATATCTCCTCTATC	rev primer to clone RyeG ORF3 into pCRT7; HindIII
JVO-16907	P-CAATGATGCATTTAAGTAATCGTCTGC	fwd primer to introduce stop codon into RyeG ORF2 on plasmids; phosphorylated; use with JVO-16550

Table 5.11, *continued.*

Name	Sequence (5'→3')	Purpose
JVO-16908	GTTTTGGTACCGTGGGCAATGATGCATTTAAG	fwd primer to clone RyeG ORF2 into pZE12/pCRT7; KpnI
JVO-16909	GTTTTTCTAGACTATAATACCAACCTAATTTCTCC	rev primer to clone RyeG ORF2 into pZE12; XbaI
JVO-16910	GTTTTAAGCTTCTATAATACCAACCTAATTTCTCC	rev primer to clone RyeG ORF2 into pCRT7; HindIII
JVO-16912	GCCACCTGATTGGGTGGCTTC	rev primer for in vitro transcription of shortened csRNA1; use with JVO-15297
JVO-16913	GCCACCGAATGCGGTGACTC	rev primer for in vitro transcription of shortened csRNA3; use with JVO-14876
JVO-16914	GCCACCCATACAGGCGAC	rev primer for in vitro transcription of shortened csRNA5; use with JVO-14880
JVO-16915	GCACCTAGTTTCCTAGATGCTAGC	rev primer for in vitro transcription of shortened F41; use with JVO-14888
JVO-16946	CCCACGGTACCTTTCTCCTC	Rev primer to mutate RyeG on pCRT7; use with JVO-16871/16907
JVO-17434	CCTGATTGGGTGGCTTCATT	NB probe for S. pneumoniae csRNA1
JVO-17435	CGGTGGAGTTAAGGGAGATTA	NB probe for S. pneumoniae csRNA2
JVO-17946	TAAAGTTAGGAGGTCTTTATTTAAT	NB probe 5' end of S. pneumoniae csRNA1/2
JVO-17953	GAAATCCTCATCTCCCCACC	NB probe S. pneumoniae comC
M13 fwd	GTAAAACGACGGCCAG	plasmid sequencing
M13 rev	CAGGAAACAGCTATGAC	plasmid sequencing
pZE-A	GTGCCACCTGACGTCTAAGA	plasmid sequencing
pZE-Xba	TCGTTTTATTTGATGCCTCTAGA	plasmid sequencing
RNA34	GGGAGACCUAGCCU	Negative ctrl for pull-downs
RNA35	AGGCUAGGUCUCCC-Biotin	2'-O-methyl-RNA oligo adaptor with 3' biotin tag

5.4 General methods

All solutions, media and equipment mentioned in this section are listed in 5.1 and 5.2. For work with RNA, only DEPC-treated water was used.

5.4.1 *Escherichia coli*

5.4.1.1 Culture

E. coli was grown on LB agar plates at 37 °C. Single colonies were then inoculated in 2 ml LB and grown overnight at 37 °C with shaking at 220 rpm. 1:100 dilutions of the overnight cultures were used in order to start the main culture, which was grown in LB at 37 °C and 220 rpm. If necessary, the following antibiotics were added to the plates or the liquid medium:

- Ampicillin/carbenicillin (Amp/Carb): 100 µg/ml

- Chloramphenicol (Cm): 20 µg/ml

- Kanamycin (Km): 50 µg/ml

5.4.1.2 Competent cells and transformation

For transformation, electrocompetent cells were prepared by growing 100 ml cells to an $OD_{600\,nm}$ of 0.4, which were then cooled to 4 °C and washed once with 15 ml ice-cold water, followed by two washes with 10 ml of ice-cold 10 % glycerol. The pellet was resuspended in 0.5 ml ice-cold 10 % glycerol and split into 80 µl aliquots that were either used directly for transformation or were stored at –80 °C.

For transformation, one aliquot was thawed on ice. 1 µl of a plasmid or PCR product was added to a pre-chilled 0.1 cm cuvette and the electrocompetent cells added. After incubation for 30–60 s on ice, an electrical pulse (1.8 kV, 200 Ω, 25 µF) was used to electroporate the cells. Recovery of the cells was performed by addition of 1 ml of 37 °C LB for 1–2 h at 37 °C and 50 rpm. The cells were then plated on LB agar plates.

5.4.1.3   Gene inactivation

Gene inactivation mostly followed a published protocol (Datsenko and Wanner, 2000). Briefly, a strain carrying the pKD46 plasmid, which carries the λRED recombinase and is temperature-sensitive, was grown overnight at 28 °C. The next day, the overnight culture was diluted 1:300 in 50 ml LB containing 0.2 % L-arabinose and grown at 28 °C to an $OD_{600\,nm}$ of 0.5. Electrocompetent cells were prepared and transformed with 800–1,000 ng of a gel-purified PCR product containing a kanamycin resistance cassette. The PCR product was obtained using the pKD4 plasmid and primers containing the flanking regions of the gene of interest. The transformed cells were streaked on LB agar plates and incubated at 37 °C overnight. The deletion was verified by PCR.

5.4.1.4   3xFLAG-tagging of genes

3xFLAG-tagging of genes mostly followed a published protocol (Uzzau et al., 2001). The steps are the same as described for the gene inactivation, except that the PCR product was obtained using the pSUB11 plasmid.

5.4.1.5   P1 transduction

In order to transduce gene knockouts or 3xFLAG-tagged genes to a recipient strain, transduction with phage P1 was used. To obtain a phage lysate, an overnight culture of the strain carrying the mutation was diluted in 10 ml 5 mM $CaCl_2$ and grown for 30 min. 200 µl of a P1 lysate obtained from a wild-type strain was added and the culture grown for 2–3 h. Lysis was stopped with 300 µl chloroform, followed by vortexing and storage in a glass tube at 4 °C.

Transduction was performed using 1 ml of an overnight culture of the recipient strain, which was resuspended in 500 µl P1 buffer. 5, 20 or 40 µl P1 lysate was added and incubated for 30 min at 37 °C without shaking. Lysis was stopped with 1 ml LB containing 0.1 M sodium citrate. After growth at 37 °C and 220 rpm, the cells were collected in 100 µl LB, streaked on LB agar plates containing 5 mM sodium citrate and grown at 37 °C. Single colonies were re-streaked twice on LB agar plates containing 5 mM sodium citrate and verified by PCR.

5.4.1.6 Removal of antibiotics resistance

To remove antibiotics resistance cassettes introduced by knockouts or 3xFLAG-tagging, a published protocol was mostly followed (Datsenko and Wanner, 2000). Briefly, the mutant strain was made electrocompetent and transformed with the temperature-sensitive pCP20 plasmid that carries the Flp recombinase. Recovery and overnight growth on plates were performed at 28 °C. Single colonies were streaked on three plates containing either no antibiotics, only Amp/Carb or only Km. The cured strain only grows on the plate without antibiotics and was additionally verified by PCR.

5.4.1.7 Cloning

To obtain plasmids with the desired inserts, inserts were generated by PCR with flanking restriction sites. Restriction digestion on the insert and the plasmid was then performed using the appropriate enzymes and the manufacturer's instructions. 75 ng of the gel-purified insert was then ligated with 25 ng of the linearized plasmid using 1 U T4 DNA ligase for $\geq$ 1 h at room temperature in a 10 μl reaction. 1 μl of the ligation reaction was used to transform electrocompetent *E. coli* cells.

5.4.1.8 Growth curves

Growth curves were performed by scraping an agar plate of the strain of interest and diluting it to a starting $OD_{600 nm}$ of 0.005 in a transparent 96 well plate. Growth and measurements were performed using a microplate reader. Alternatively, the dilution was performed in 50 ml LB and the measurements performed manually.

5.4.2 *Streptococcus pneumoniae*

5.4.2.1 Culture

S. pneumoniae was grown on tryptic soy agar containing 5 % sheep blood (TSA-B) plates at 37 °C and 5 % CO_2. Pre-cultures were grown by scraping whole plates and diluting the bacteria in pre-warmed THY and growing them to an $OD_{600 nm}$ of 0.5

at 37 °C without shaking. The main culture was then prepared by refreshing the
pre-culture in THY to a starting $OD_{600\,nm}$ of 0.5 and grown at 37 °C without shaking.

5.4.2.2   Gene inactivation and 3xFLAG-tagging

Knockout and 3xFLAG-tagging were performed by homologous recombination
using a PCR product containing a spectinomycin cassette that was flanked by
$\sim$1,000 bp of the gene of interest. The PCR product was obtained by overlapping
PCR of the up- and downstream regions of the gene of interest as well as a specti-
nomycin resistance cassette amplified from pSP72::Sp. To obtain competent cells,
a wild-type strain was grown to an $OD_{620\,nm}$ of 0.5 at 37 °C without shaking in
C+Y (Lacks and Hotchkiss, 1960), pH 8. This pre-culture was refreshed in the same
medium to a starting $OD_{620\,nm}$ of 0.05 and grown to an $OD_{620\,nm}$ of 0.13 at 37 °C
without shaking. 1:10 dilutions in pre-warmed C+Y, pH 8 were incubated for 30 min
at 30 °C, followed by addition of CSP-1 (for strain R6) or CSP-2 (for strain TIGR4) to
a concentration of 100 ng/ml for 15 min. 100 ng/ml of the PCR product was added,
followed by incubation for 60 min at 30 °C and 90 min at 37 °C. Cells were streaked
on TSA-B plates containing 200 µg/ml spectinomycin and grown at 37 °C and 5 %
CO_2. Verification of the mutants was performed by PCR.

5.4.2.3   Growth curves

Growth curves were performed as described for *E. coli*.

5.4.3   DNA and RNA methods

5.4.3.1   Polymerase chain reaction (PCR)

To amplify DNA, PCR was performed using the *Taq* or Phusion DNA polymerases
and purification was performed using the NucleoSpin Gel and PCR clean-up kit
following the manufacturer's instructions. Oligonucleotides used are listed in Table
5.11.

5.4.3.2 Agarose gel electrophoresis

To separate DNA, 5 volumes of sample were mixed with 1 volume of $6\times$ DNA loading buffer and run using 0.8–2% (w/v) agarose gels in $1\times$ TAE for 30–90 min at 100–160 V. Gels were stained in an ethidium bromide bath for 15 min before visualization.

5.4.3.3 Hot phenol RNA extraction

To isolate total RNA from bacteria, $4 \, \mathrm{OD}_{600\,\mathrm{nm}}$ of cells was mixed with 0.2 volumes of ice-cold stop mix followed by freezing in liquid nitrogen. Samples were thawed on ice and the cells pelleted. For *E. coli*, 600 µl of $1\times$ TE with 0.5 mg/ml lysozyme were added. For *S. pneumoniae*, 600 µl of $1\times$ TE with 10 mg/ml lysozyme were added followed by incubation at 37 °C for 10 min. The rest of the protocol did not differ between the organisms: 60 µl of 10 % (w/v) SDS was added and the solution incubated for 1–2 min at 64 °C. 66 µl 3 M NaOAc, pH 5.2 and 750 µl acidic phenol were added and incubated for 6 min at 64 °C. After cooling the tubes on ice, phases were separated for 15 min at 16,100 g and 4 °C. The aqueous phase was transferred to a PLG tube and 750 µl chloroform was added. Phases were separated for 15 min at 16,100 g and 4 °C. The aqueous layer was transferred to a new tube and 1.4 ml of 30:1 mix, pH 6.5 was added to precipitate the RNA for $\geq$1 h at –20 °C. The RNA was collected by centrifugation for 30 min at 16,100 g and 4 °C. The pellet was washed with 350 µl 70 % ethanol followed by centrifugation for 10 min at 16,100 g and 4 °C. The purified RNA was then diluted in water and stored at –20 °C.

5.4.3.4 Phenol/chloroform/isoamyl alcohol (P/C/I) RNA extraction

RNA extraction using P/C/I was performed by adding 1 volume of acidic P/C/I to the sample. Phases were separated for 15 min at 16,100 g and 4 °C. The aqueous layer was transferred to a new tube and 3 volumes of 30:1 mix, pH 6.5 were added to precipitate the RNA for $\geq$1 h at –20 °C. If the RNA concentration was expected to be low, 1 µl GlycoBlue was added as co-precipitant. The RNA was collected by centrifugation for 30 min at 16,100 g and 4 °C. The pellet was washed with 350 µl

70 % ethanol followed by centrifugation for 10 min at 16,100 g and 4 °C. The purified RNA was then diluted in water and stored at –20 °C.

5.4.3.5 DNase I digestion of RNA

To remove DNA from an RNA sample, DNA was digested by addition of 1 U of DNase I per mg of RNA and incubation at 37 °C for 45 min. The DNase-treated RNA was then purified using P/C/I extraction.

5.4.3.6 Denaturing polyacrylamide gel electrophoresis (PAGE)

To separate RNA, 1 volume of sample was mixed with 1 volume of 2× RNA loading buffer, incubated for 5 min at 95 °C and then put on ice. The samples were then run using 6–10 % polyacrylamide gels with 7 M urea in 1× TBE for 90–150 min. Gels were either stained in an ethidium bromide bath for 15 min before visualization or blotted.

5.4.3.7 Northern blotting

After RNA PAGE, gels were transferred onto Hybond+ membranes in 1× TBE at 50 V for 1 h using a wet blotting system. The RNA was crosslinked to the membrane at 0.12 J and 254 nm followed by pre-hybridization in 15 ml Roti Hybri-Quick for 1 h at 42 °C. 2–5 pmol of a radioactively labeled DNA oligonucleotide was added and hybridized at 42 °C overnight. The membrane was washed at 42 °C for 15 min each with 5× SSC-S, 1× SSC-S and 0.5× SSC-S. Then, the membrane was dried, sealed in plastic foil and exposed on a phosphor screen.

10 pmol labeled DNA oligonucleotides were generated by phosphorylation using 0.5 µl T4 polynucleotide kinase, 10 µCi ^{32}P-γ-ATP in a 10 µl reaction for 45 min at 37 °C. The reaction was purified from unincorporated nucleotides by centrifugation using G-25 columns.

5.4.3.8 *In vitro* transcription

To synthesize RNA *in vitro*, the sequence of interest was amplified with a T7 promoter by PCR. Transcription was performed using the TranscriptAid kit following

the manufacturer's instructions. 40 µl transcription reaction were mixed with 40 µl 2× RNA loading buffer and separated using 6 % denaturing PAGE. The gel was stained with ethidium bromide and the band of expected size was excised. RNA was extracted from the gel by addition of 750 µl of RNA extraction buffer followed by rotating overnight at 4 °C. The supernatant was then subjected to P/C/I extraction and the size of the RNA checked on denaturing PAGE.

5.4.4 Protein methods

5.4.4.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

To separate proteins, cell pellets were dissolved in 100 µl of 1× protein loading buffer. Alternatively, 4 volumes of sample were mixed with 1 volume of 5× protein loading buffer. The samples were then incubated for 5 min at 95 °C and kept at room temperature. Separation was performed using 12 or 15 % SDS polyacrylamide gels in 1× running buffer for 1.5–3 h at 45 mA. Gels were then either stained using Coomassie or blotted.

5.4.4.2 Western blotting

After SDS-PAGE, gels were transferred to PVDF membranes in 1× transfer buffer at 340 mA for 1.5 h using a semidry blotting system. Prior to blotting, the membranes were activated by incubation for 90 s in methanol, 5 min in water and 5 min in 1× transfer buffer. After blotting, the membranes were blocked in 5 % powdered milk in 1× TBS-T for 1 h at room temperature. Following 3 washes for 2 min with 1× TBS-T, the appropriate primary antibody was added (diluted in 1× TBS-T and 3 % BSA) and incubated overnight at 4 °C. The membranes were then washed three times for 15 min with 1× TBS-T and the appropriate secondary antibody (diluted in 1× TBS-T) was added and incubated for 1 h at room temperature. Finally, the membranes were washed twice for 10 min with 1× TBS-T and then developed using chemiluminescence detection solution.

5.5 Grad-seq

All solutions, media and equipment mentioned in this section are listed in 5.1 and 5.2. The Grad-seq method described here is based on a previous study in *Salmonella* (Smirnov et al., 2016) and has been published based on its application to *S. pneumoniae* (Hör et al., 2020a). For all steps, only DEPC-treated water was used.

5.5.1 Preparation of glycerol gradients

The middle of a clean open-top polyallomer tube was marked. The 10 % glycerol solution was filled in so it exceeded the mark at the middle of the tube by ∼2–3 mm. Using a 70 mm injection needle and a syringe, the 40 % glycerol solution was then injected underneath the 10 % glycerol solution until the interphase reached exactly the mark at the middle of the tube. The tube was closed with a short cap and the liquid inside of the cap carefully removed. Using the Gradient Master Station, a linear 10–40 % glycerol gradient was prepared with the 10–40 % (w/v) glycerol short cap program. The prepared gradients were stored at 4 °C until needed.

5.5.2 Preparation of lysates

5.5.2.1 *Escherichia coli*

E. coli K-12 MG1655 was grown in 100 ml LB to an $OD_{600\,nm}$ of 2.0 (early stationary phase). Following cooling in an ice bath, the cells were washed three times with ice-cold 1× TBS and collected in 500 µl lysis buffer A (+ 20 U/ml DNase I, 200 U/ml RNase inhibitor and 0.2 % (v/v) Triton X 100), followed by addition of 750 µl 0.1 mm glass beads. Mechanical lysis was performed by vortexing (highest power) for 30 s followed by 15 s cooling on ice. This process was repeated ten times. Finally, the lysate was cleared by centrifugation for 10 min at 16,100 g and 4 °C.

5.5.2.2 *Streptococcus pneumoniae*

S. pneumoniae TIGR4 was grown in 400 ml THY to an $OD_{600\,nm}$ of 0.5 (mid-logarithmic phase). Following cooling in an ice bath, the cells were washed three times with ice-cold 1× TBS, collected in 500 µl lysis buffer A (+ 20 U/ml DNase

I, 200 U/ml RNase inhibitor and 0.2 % (v/v) Triton X 100) and transferred to a 2 ml FastPrep tube with lysing matrix E. Mechanical lysis was performed using the FastPrep-24 instrument at 6 m/s for 30 s. Finally, the lysate was cleared by centrifugation for 30 min at 16,100 g and 4 °C.

5.5.3 Gradient centrifugation and fractionation

10 µl of the cleared lysate was mixed with 1 ml TRIzol for the RNA input control and 20 µl were mixed with 20 µl 5× protein loading buffer for the protein input control. Then, 200 µl (*E. coli*) or 400 µl (*S. pneumoniae*) lysate were layered on top of the 10–40 % glycerol gradient after the same volume was removed from the gradient to prevent spilling. Gradient centrifugation was performed for 17 h at 4 °C and 100,000 g (23,700 rpm) using an SW 40 Ti rotor. Acceleration and deceleration were both set to 9.

Fractionation of the gradient was performed manually by pipetting 20 590 µl fractions from the top of the gradient using a P-1,000. The pellet was collected by resuspension in the remaining ∼300 µl. To obtain a UV profile of the gradient, the A$_{260\,nm}$ values of each fraction were measured. 90 µl of each fraction were mixed with 30 µl 5× protein loading buffer to obtain the samples for protein analysis and stored at –20 °C.

5.5.4 RNA extraction

50 µl of 10 % SDS (25 µl for the pellet) and 600 µl acidic P/C/I (300 µl for the pellet) were added to the remaining 500 µl of each fraction. To ensure dissociation of RNPs, each fraction was vortexed for 30 s and let rest at room temperature for 5 min. Phases were separated by centrifugation for 15 min at 16,100 g and 4 °C. The aqueous phases were collected and precipitated by addition of 1 µl GlycoBlue and 1.4 ml of ice-cold 30:1 mix, pH 6.5 and incubation for ≥1 h at –20 °C. The RNA was collected by centrifugation for 30 min at 16,100 g and 4 °C. The pellet was washed with 350 µl of 70 % ethanol followed by centrifugation for 10 min at 16,100 g and 4 °C. The lysate input control stored in TRIzol was extracted according to the manufacturer's instructions, except that 400 µl chloroform was used, 1 µl GlycoBlue was added and that precipitation was performed using 30:1 mix, pH 6.5.

The purified RNA was dissolved in 40 μl water and 5 μl 10× DNase I buffer, 4 μl DNase I, 0.5 μl RNase inhibitor and 0.5 μl water were added. DNA was then digested by incubation for 45 min at 37 °C. 150 μl water and 200 μl acidic P/C/I were added and P/C/I extraction performed as above. The final RNA pellet was diluted in 35 μl water and stored at –80 °C.

5.5.5 Northern and western blotting

For northern and western blotting, equal volumes of each fraction were loaded on the gels to keep the information of gradient separation.

5.5.6 RNA-seq

5.5.6.1 Library preparation and sequencing

For RNA-seq, 5 μl purified RNA from each gradient fraction plus the pellet were diluted in 45 μl water. Of this dilution, 10 μl were mixed with 10 μl of a 1:100 dilution of the ERCC spike-in mix 2. Library preparation and sequencing was performed by Vertis Biotechnologie AG. Briefly, the RNA was fragmented using ultrasound (4 pulses of 30 s at 4 °C) and the 3′ adapter was ligated. First strand cDNA synthesis was carried out using M-MLV reverse transcriptase. Then, the 5′ Illumina TruSeq sequencing adapter was ligated to the 3′ end of the antisense cDNA, followed by PCR amplification to 10–20 ng/μl. The PCR reaction was purified and pooled with ratios according to RNA concentration of the fractions. A preparative agarose gel was run to size-select a range of 200–550 bp for the cDNA, which was subsequently sequenced on an Illumina NextSeq 500 instrument using 75 nt single-end read length.

5.5.6.2 RNA-seq analysis

Sequencing reads were trimmed and clipped using cutadapt (Martin, 2011). Read filtering, read mapping, nucleotide-wise coverage calculation and genome feature-wise counting were performed using READemption (Förstner et al., 2014) and segemehl (Hoffmann et al., 2014). The genome versions used were NC_000913.3 for *E. coli* MG1655 and NC_003028.3 for *S. pneumoniae* TIGR4. ERCC spike-in-based normaliza-

tion and further analyses were performed using the tool GRADitude[1] (Di Giorgio,
Hör, Vogel and Förstner, unpublished). These steps were performed by Silvia Di
Giorgio (ZB MED, Cologne). Based on quantified northern blots, the normalized
data was manually adjusted to remove left-over disturbances in the data: for *E. coli*,
fractions 5, 7 and 8 were multiplied by 1.5, 4.5 and 28, respectively (see 2.4.3); for *S.
pneumoniae*, fractions 3, 4, 5, 6, 7 and 15 were multiplied by 1.15, 1.15, 0.85, 0.95, 0.85
and 1.25, respectively.

5.5.7 Mass spectrometry

5.5.7.1 Sample preparation and MS

Gradient protein samples were homogenized by ultrasound using a Bioruptor Plus
instrument (5 cycles of 30 s on followed by 30 s off, high power, at 4 °C), followed
by centrifugation for 15 min at 16,100 g and 4 °C. 20 µl of the cleared protein sample
was mixed with 10 µl UPS2 spike-in (diluted in 250 µl 1.25× protein loading buffer).
Next, 50 mM DTT were added and the samples were reduced for 10 min at 70 °C,
followed by alkylation with 120 mM iodoacetamide for 20 min at room temperature
in the dark. The samples were precipitated by addition of four volumes of acetone
and incubation at –20 °C overnight. After four washes with acetone at –20 °C, the
pellets were dissolved in 50 µl 8 M urea and 100 mM ammonium bicarbonate, pH 8.

0.25 µg Lys-C were added to digest the proteins for 2 h at 30 °C. Subsequent
dilution to 2 M urea by addition of 150 µl 100 mM ammonium bicarbonate, pH 8
was followed by overnight digestion with 0.25 µg trypsin at 37 °C. C-18 Stage Tips
(Rappsilber et al., 2003) were used to desalt the peptides, which were then eluted
with 60 % acetonitrile and 0.3 % formic acid. Finally, the peptides were dissolved in
2 % acetonitrile and 0.1 % formic acid and subjected to nanoLC-MS/MS.

5.5.7.2 NanoLC-MS/MS analysis

Nano liquid chromatography followed by tandem mass spectrometry (nanoLC-
MS/MS) was performed by the lab of Andreas Schlosser at the Rudolf Virchow
Center in Würzburg using an Orbitrap Fusion instrument equipped with a PicoView

---

[1]https://foerstner-lab.github.io/GRADitude/

ion source coupled to an EASY-nLC 1000 and similar to Cossa et al. (2020). The peptides were separated with a 140 min linear gradient from 3 to 40 % acetonitrile and 0.1 % formic acid at a flow rate of 500 nl/min. Both MS and MS/MS scans were acquired in the Orbitrap analyzer with a resolution of 60,000 for MS scans and 15,000 for MS/MS scans. HCD fragmentation with 35 % normalized collision energy was applied. A top speed data-dependent MS/MS method with a fixed cycle time of 3 s was used. Dynamic exclusion was applied with a repeat count of 1 and an exclusion duration of 60 s; singly charged precursors were excluded from selection. Minimum signal threshold for precursor selection was set to 50,000. Predictive AGC was used with a target value of 2×10^5 for MS scans and 5×10^4 for MS/MS scans. EASY-IC was used for internal calibration.

Raw MS data files were analyzed with MaxQuant version 1.5.7.4 (Cox and Mann, 2008). Database search was performed using Andromeda (integrated into MaxQuant) against the UniProt database for *E. coli* MG1655 (UP000000625, organism identifier: ECOLI) or *S. pneumoniae* TIGR4 (UP000000585, organism identifier: STRPN), a database containing the UPS2 spike-in and a database containing common contaminants. The search was performed with tryptic cleavage specificity with 3 allowed miscleavages. A false-discovery rate of 1 % on both protein and peptide level controlled protein identification. In addition to the MaxQuant default settings, the search was performed against the following variable modifications: protein N-terminal acetylation, Gln to pyro-Glu formation (N-terminal Gln) and oxidation of Met. For protein quantitation, the LFQ intensities were used (Cox et al., 2014). Proteins with less than 2 identified razor/unique peptides were dismissed.

Normalization was performed by Silvia Di Giorgio (ZB MED, Cologne) based on the UPS2 spike-in using the tool GRADitude[2] (Di Giorgio, Hör, Vogel and Förstner, unpublished).

5.5.7.3 Data availability

The *E. coli* and *S. pneumoniae* Grad-seq datasets can be retrieved from a USB thumb drive attached to the back of this thesis. The *S. pneumoniae* Grad-seq data have further been published recently (Hör et al., 2020a).

---

[2]https://foerstner-lab.github.io/GRADitude/

5.6 Methods used in chapter 2

This section describes the specific methods used in chapter 2, *Grad-seq of* Escherichia coli. All solutions, media, equipment and bacterial strains mentioned in this section are listed in 5.1, 5.2 and 5.3. All used oligonucleotides are listed in Table 5.11. For work with RNA, only DEPC-treated water was used.

5.6.1 Purification of ribosomes

Crude purification of ribosomes mostly followed a previously published protocol (Mehta et al., 2012). Briefly, 800 ml of an *E. coli* culture was grown to an $OD_{600\,nm}$ of $\sim$0.5–0.7 and washed once with 25 ml of ice-cold 1$\times$ TBS. The cell pellets were snap-frozen in liquid nitrogen and stored at –80 °C. The pellets were resuspended in 6 ml ice-cold lysis buffer B on ice. Lysis was performed by two lysis steps using a french press at 10,000 psi. 75 µl of 100 mM PMSF was added and the lysates cleared by centrifugation for 30 min at 4 °C and 30,000 g using an SW 40 Ti rotor. 12.5 ml of the supernatant was subsequently layered on top of a 12.5 ml 1.1 M sucrose cushion made up in lysis buffer B. Next, the sample was centrifuged for 16 h at 4 °C and 100,000 g using a type 70 Ti rotor.

The pellet was gently washed with 500 µl of storage buffer and finally dissolved in 1 ml of storage buffer by gentle shaking for 2.5 h at 4 °C. After centrifugation for 5 min at 16,100 g and 4 °C, the concentration was measured, the purified ribosomes aliquoted, snap-frozen in liquid nitrogen and stored at –80 °C.

5.6.2 Sucrose polysome gradient analysis

5.6.2.1 Analysis of total lysates

50 ml of an *E. coli* culture was grown to an $OD_{600\,nm}$ of 2, followed by rapid filtration and immediate freezing in liquid nitrogen. The cells were then thawed on ice, resuspended in 1 ml of ice-cold lysis buffer C (+ 1 mM PMSF, 0.4 % Triton X 100, 20 U/ml DNase I and 200 U/ml RNase-inhibitor), transferred to a 2 ml FastPrep tube with lysis matrix E and lysed using a FastPrep-24 instrument for 15 s at 4 m/s. The lysate was cleared by centrifugation for 10 min at 4 °C and 16,100 g. Of the cleared

lysate, 10 µl was collected as input control. An equivalent of 15 $A_{260\,nm}$ per ml of the cleared lysate was layered on top of a linear 10–55 % (w/v) sucrose gradient (in lysis buffer C + 5 mM $CaCl_2$), which was formed in an open-top polyclear tube (see 5.5.1). The gradient was centrifuged for 2.5 h at 4 °C and 237,000 g (35,000 rpm) using an SW 40 Ti rotor. Acceleration and deceleration were both set to 9. Subsequently, the gradient was automatically fractionated into 20 fractions using a fractionator. Simultaneously, a UV profile at 254 nm was recorded.

When the goal was to analyze RNA from the gradient, RNA extraction was performed as for glycerol gradients (see 5.5.4), except that the vortexing step was performed for 15 s and that DNase treatment was skipped. Northern blotting was performed as for glycerol gradients (see 5.5.5).

When the goal was to analyze protein from the gradient, 4 volumes of sample were mixed with 1 volume of 5× protein loading buffer and subjected to western blotting as performed for glycerol gradients (see 5.5.5). If necessary, the protein content of each fraction was precipitated by addition of 1/10 volume of ice-cold 100 % trichloroacetic acid and 1/10 volume of 0.15 % desoxycholate. The samples were briefly shaken and let rest on ice for 10 min. After centrifugation at 16,100 g and 4 °C for 10 min, the pellet was washed once with 400 µl ice-cold acetone and dissolved in 50 µl of 5× protein loading buffer diluted to 1× with 1 M Tris-HCl, pH 8.6.

5.6.2.2   Analysis of *in vitro*-reconstituted complexes

To test binding of RyeG or YggL to purified ribosomes (see 5.6.1), given amounts of *in vitro*-transcribed RyeG or recombinant YggL (produced by the Recombinant Protein Expression core unit at the Rudolf Virchow Center in Würzburg) were mixed with purified ribosomes extracted from the corresponding knockout strain. Then, the volume was increased to 200 µl lysis buffer B and the samples were incubated for 10 min at 30 °C with shaking at 330 rpm to allow complex formation. The samples were subsequently loaded on 10–40 % (w/v) sucrose gradients (in lysis buffer D), which were formed in open-top polyclear tubes (see 5.5.1). Gradients were centrifuged for 14 h at 4 °C and 71,000 g (20,000 rpm) using an SW 40 Ti rotor. Acceleration and deceleration were both set to 9. Fractionation as well as RNA and protein analysis were performed as described in 5.6.2.1.

5.6.3 Estimation of *in vivo* RNA copy numbers

Estimation of *in vivo* copy numbers of RyeG was performed as described previously (Fröhlich et al., 2013). Briefly, RNA was extracted at the given time points by collecting 4 $OD_{600\,nm}$ of cells. The RNA was diluted in 40 µl water and 10 µl of each time point ($\approx 10^9$ cells) was subjected to northern blotting. As reference, *in vitro*-transcribed RyeG was loaded (0.05, 0.1, 0.5, 1 and 2.5 ng). RNA levels per cell were based on determination of viable cell counts per $OD_{600\,nm}$ as described in Sittka et al. (2007).

5.6.4 Microscopy

Strains were grown on LB agar plates overnight. Plates were scraped and cultures were inoculated in LB at a starting $OD_{600\,nm}$ of 0.05. After 1.5 h of growth, 1 ml of cells was collected and washed once with 1× TBS. Next, the cells were fixed with 500 µl of 4 % paraformaldehyde for 15 min in the dark. 4 µl were spotted on a cover slip covered with 1.5 % agarose in 1× PBS to immobilize the cells and imaged using a confocal microscope.

5.6.5 30S subunit toeprinting analysis

30S subunit toeprinting was performed as previously published (Hartz et al., 1988; Smirnov et al., 2017b) with few changes. Briefly, 0.2 pmol unlabeled, *in vitro*-transcribed RyeG and 0.5 pmol of a 5'-labeled DNA oligonucleotide (JVO-16833) were denatured for 1 min at 95 °C in the presence of 0.8 µl SB 5× –Mg in a total volume of 3 µl. After incubation on ice for 5 min, 1 µl dNTPs (5 mM each) and 1 µl SB 1× Mg60 were added and the samples were incubated for 5 min at 37 °C. Next, 4 pmol purified 30S subunits (pre-activated for 20 min at 37 °C) was added to the samples (SB 1× Mg10 was added to the control). After incubation for 5 min at 37 °C, 10 pmol uncharged fMet-tRNA$_i^{Met}$ was added to the corresponding sample.

Reactions were continued at 37 °C for 15 min, followed by addition of 100 U SuperScript II reverse transcriptase and incubation for 20 min at 37 °C. Reactions were stopped by addition of 100 µl toeprint stop buffer. DNA was extracted by addition of 110 µl P/C/I and extraction. Next, 5 µl 3 M KOH was added and the RNA

digested at 90 °C for 5 min. 10 µl 3 M acetic acid, 1 µl GlyoBlue and 300 µl 30:1 mix, pH 6.5 were added and the DNA precipitated at –20 °C overnight. Extraction was finished and the pellet washed once with 100 µl of 70 % ethanol. The purified pellet was dissolved in 10 µl 1× RNA loading buffer, denatured for 3 min at 90 °C and subjected to separation using a denaturing 8 % sequencing gel in presence of a RyeG-specific sequencing ladder prepared using the DNA Cycle Sequencing kit according to the manufacturer's instructions. Gels were run for 1.5 h at 40 W, dried and exposed on a phosphor screen.

5.6.6   Affinity purification followed by MS (AP/MS)

50 $OD_{600\,nm}$ of the *yggL*-3xFLAG and wild-type strains were collected and washed once with 1 ml of lysis buffer A. After resuspension in 800 µl lysis buffer A, the cells were transferred to a 2 ml FastPrep tube with lysis matrix E and lysed using a FastPrep-24 instrument for 20 s at 4 m/s. The lysate was cleared for 10 min at 16,100 g and 4 °C.

40 µl of magnetic protein A/G beads were washed with 1 ml of lysis buffer A, resuspended in 400 µl lysis buffer A and 3 µl anti-FLAG antibody was added. After rotating for 45 min at 4 °C, the beads were washed twice with 400 µl lysis buffer A.

600 µl of the lysate was added to the beads with the coupled antibody and rotated for 1.5 h at 4 °C. The beads were washed five times with 400 µl lysis buffer A and briefly spun down. The lysis buffer was removed, the beads were resuspended in 35 µl 1× LDS sample buffer with 50 mM DTT and the proteins eluted by incubation at 95 °C for 5 min. Alkylation, gel electrophoresis and MS were performed as described in 5.7.2.

5.7   Methods used in chapter 3

This section describes the specific methods used in chapter 3, *Grad-seq of* Streptococcus pneumoniae. Most of these methods have been published previously in the corresponding paper (Hör et al., 2020a). All solutions, media, equipment and bacterial strains mentioned in this section are listed in 5.1, 5.2 and 5.3. All used

oligonucleotides are listed in Table 5.11. For work with RNA, only DEPC-treated water was used.

5.7.1　MS2 pull-down and silver staining

MS2 pull-downs with *in vitro*-transcribed ncRNAs were performed as previously described (Said et al., 2009; Smirnov et al., 2016), with some modifications. Briefly, 200 OD$_{600\,nm}$ of *S. pneumoniae* TIGR4 wild type grown to mid-logarithmic phase were used for each pull-down. The templates for the *in vitro*-transcribed bait ncRNAs were created by overlapping PCR with ncRNA- and MS2-specific primers. Both 5'- and 3'-located MS2 tags were tried. Changes to most of the steps of the protocol were tested in order to achieve pull-down of proteins: lysis conditions, number of cells per pull-down, number of washes, MS2-MBP stocks, concentrations of bait RNAs and MS2-MBP, folding of the bait RNAs, amount of amylose resin and more. 5'-MS2-tagged ChiX was used as a positive control with *Salmonella* lysates, which achieved pull-down of Hfq, indicating that a pneumococcus-specific issue prevented recovery of proteins.

Silver staining of gels was performed using 15 % SDS-PAGE to separate the proteins, followed by incubation in fixing solution for 1–3 h. Then, the gel was washed twice in 50 % ethanol for 20 min, incubated for 1 min in sensitizer and washed three times for 20 s with water. After incubation in silver staining solution for 20 min and two 20 s washes with water, the gel was developed using developer solution for 5–30 min. The process was stopped by incubation in silver stop solution for 2 min and the gel rinsed with water.

5.7.2　14mer pull-down

5.7.2.1　14mer pull-down assay

The 14mer pull-down used in this thesis (see 3.4.2) is based on a previously published method (Treiber et al., 2017, 2018). To obtain *in vitro*-transcribed RNAs carrying the 14mer tag, PCR templates were generated with a 39 nt 5' overhang: GTTTTTTT*TAATACGACTCACTATA*__GGG__**AGACCTAGCCT**, where highlighted nucleotides represent the *T7 promoter*, the **14mer tag** and the <u>TSS</u>. The negative control

for the pull-down experiments was an RNA oligonucleotide only containing the 14mer tag (GGGAGACCUAGCCU).

For the pull-down, 100 µl magnetic streptavidin beads was washed three times with 1 ml of lysis buffer E. The rest of the protocol was performed at 4 °C. 4 µg of a 3′-biotinylated, 2′-O-methyl-modified RNA adaptor complementary to the 14mer tag of the bait RNAs (AGGCUAGGUCUCCC-biotin) was coupled to the washed beads for 1 h with rotation. Following two washes with 1 ml of lysis buffer E, the adaptor-coupled beads were resuspended in 1 ml of lysis buffer E. Two enable pre-clearing of the lysate, the beads were split into two tubes with 500 µl each. One of the tubes was used to couple 10 µg per 100 nt of bait RNA rotating overnight, the other was stored.

100 $OD_{600\,nm}$ of *S. pneumoniae* TIGR4 wild type was lysed and cleared as described in 5.5.2.2, except that the cells were resuspended in 500 µl lysis buffer F. Pre-clearing of the lysate was performed by incubation with the stored beads for 3.5 h with rotation. The beads were subsequently removed by centrifugation for 10 min at 16,100 g. Next, the bait RNA-coupled beads were washed twice with lysis buffer E and incubated with the pre-cleared lysate supernatant for 2 h to allow the capture of interacting proteins of the bait RNAs. Finally, the beads were washed with 1 ml each of wash buffer A, wash buffer B and lysis buffer F.

To elute the captured proteins, the beads were resuspended in 35 µl of 1× LDS sample buffer containing 50 mM DTT and boiled for 5 min. The samples were alkylated in presence of 120 mM iodoacetamide for 20 min in the dark and run on a precast 4–12 % Bolt Bis-Tris plus gel using 1× MES buffer. The gel was stained with SimplyBlue Coomassie and each lane of the gel was either cut into 11 pieces or specific prominent bands were cut. To prepare the gel pieces for LC/MS-MS, they were destained with 30 % acetonitrile in 100 mM ammonium bicarbonate, pH 8. Next, the pieces were shrunk using 100 % acetonitrile and dried. Digestion was performed by addition of 0.1 µg trypsin per gel piece and incubation overnight at 37 °C in 100 mM ammonium bicarbonate, pH 8. The supernatant was removed and the peptides were extracted from the gel pieces with 5 % formic acid. Finally, the supernatant was pooled with the extracted peptides.

5.7.2.2   14mer pull-down nanoLC-MS/MS analysis

LC-MS/MS of the 14mer pull-downs was performed by the lab of Andreas Schlosser at the Rudolf Virchow Center in Würzburg similar to the protocol described in 5.5.7.2 with some exceptions: An LTQ-Orbitrap Velos Pro instrument was used. A 30 min linear gradient from 3 to 30 % acetonitrile and 0.1 % formic acid was used. MS scans were acquired in the Orbitrap analyzer with a resolution of 30,000 at m/z 400, MS/MS scans were acquired in the Orbitrap analyzer with a resolution of 7,500 at m/z 400 using HCD fragmentation with 30 % normalized collision energy. A TOP5 data-dependent MS/MS method was used; dynamic exclusion was applied with a repeat count of 1 and an exclusion duration of 30 s. Predictive AGC was used with a target value of 1×10^6 for MS scans and 5×10^4 for MS/MS scans. Lock mass option was applied for internal calibration in all runs using background ions from protonated decamethylcyclopentasiloxane.

5.7.3   CLIP-seq and CLIP-seq analysis

For CLIP-seq of Cbf1-3xFLAG, 800 ml *S. pneumoniae cbf1*-3xFLAG was grown to an $OD_{600 \, nm}$ of 0.5. Half of the culture was crosslinked, whereas the other half was used as the non-crosslinked control. The rest of the steps were performed as previously published (Holmqvist et al., 2018). Analysis of the CLIP-seq data was performed by Thorsten Bischler (Core Unit SysMed) as described previously (Chihara et al., 2019). Peaks with a log_2 fold-change of ≥ 1.0 and an adjusted *p*-value of ≤ 0.01 were considered significant.

5.7.4   *In vitro* RNase assay

400 ng of *in vitro*-transcribed ncRNAs were digested with 3.5 µg recombinant Cbf1 (produced by the Recombinant Protein Expression core unit at the Rudolf Virchow Center in Würzburg) in a total volume of 20 µl. The buffer conditions were chosen according to a previous publication (Fang et al., 2009): 50 mM Tris-HCl, pH 8, 100 mM KCl. To test the dependence on divalent cations, water, 5 mM $MgCl_2$ or 5 mM $MnCl_2$ were added. 10 % of the reaction volume was taken as samples 0, 5 and 15 min after addition of Cbf1. Immediately after the 0 min time point was taken, the reaction was

shifted to 37 °C. The samples were collected on ice in RNA loading buffer to stop the reaction and analyzed using northern blotting.

5.7.5   Rifampicin RNA stability assay

Cells were grown to an $OD_{600\,nm}$ of 0.5 and 8 ml were collected as the untreated 0 min control. Transcription was stopped with 500 µg/ml rifampicin, followed by collection of 8 ml samples after 2, 4, 8, 16 and 32 min. RNA of these samples was purified using hot phenol extraction (see 5.4.3.3) and DNase digestion (see 5.4.3.5). Of the purified RNA, 5 µg were analyzed using northern blotting.

5.7.6   CSP induction assay and agarose northern blot

Cells were grown to an $OD_{600\,nm}$ of 0.5 and 1 ml (protein sample) and 4 ml (RNA sample) of culture were collected as the untreated 0 min controls. Competence was stimulated by addition of 100 ng/ml CSP-2, whereas water was added to the control instead. 5, 10, 15 and 30 min after stimulation, further samples were collected. RNA purification was performed by hot phenol extraction (see 5.4.3.3) and DNase digestion (see 5.4.3.5). The protein samples were dissolved in 200 µl 1× protein loading buffer and 10 µl of each sample was separated using SDS-PAGE.

Of the purified RNA, 15 µg per sample was heated to 75 °C for 5 min and separated using a 1.2 % agarose gel with 1× MOPS buffer and 1.1 % formaldehyde. The gel was run in 1× MOPS buffer for 3 h at 100 V. After capillary blotting to a Hybond+ membrane overnight in 10× SSC, the protocol was continued as for standard northern blotting (see 5.4.3.7).

5.7.7   Reverse transcription-quantitative PCR (RT-qPCR)

RNA was extracted using hot phenol extraction (see 5.4.3.3) followed by DNase digestion (see 5.4.3.5). Presence of contaminating genomic DNA was tested for by PCR. RT-qPCR was carried out using the Power SYBR Green RNA-to-CT 1-Step kit and a CFX96 system. *gyrA* was used as the control gene. Data were analyzed using the comparative $\Delta\Delta C_T$ method (Livak and Schmittgen, 2001).

5.7.8 Spontaneous competence assay

For spontaneous competence assays, pre-cultures were grown to an $OD_{600\,nm}$ of 0.5 at 37 °C without shaking in C+Y, pH 8. These pre-cultures were refreshed in the same medium to a starting $OD_{600\,nm}$ of 0.05 and grown to an $OD_{600\,nm}$ of 0.13 at 37 °C without shaking. 1:10 dilutions in pre-warmed C+Y, pH 8 were incubated for 30 min at 30 °C. 100 ng/ml of a 524 bp PCR product was added, which encompassed the SmR-*rpsL* allele that carries a point mutation in the *rpsL* gene and thereby confers resistance to streptomycin (Muschiol et al., 2017). Incubation was continued for 60 min at 30 °C and 90 min at 37 °C. Cells were streaked on TSA-B plates containing 150 µg/ml streptomycin and grown at 37 °C and 5 % CO_2. Colony forming units (CFUs) were counted and compared between strains. Competence assays were performed by Geneviève Garriss (Karolinska Institutet, Stockholm, Sweden).

Chapter 6

Bibliography

Abe, Y., Fujisaki, N., Miyoshi, T., Watanabe, N., Katayama, T., and Ueda, T. (2016). Functional analysis of CedA based on its structure: residues important in binding of DNA and RNA polymerase and in the cell division regulation. *J Biochem*, 159:217–223.

Acebo, P., Martin-Galiano, A. J., Navarro, S., Zaballos, A., and Amblar, M. (2012). Identification of 88 regulatory small RNAs in the TIGR4 strain of the human pathogen Streptococcus pneumoniae. *RNA*, 18:530–546.

Akita, K., Hieda, N., Baba, N., Kawaguchi, S., Sakamoto, H., Nakanishi, Y., Yamanishi, M., Mori, K., and Toraya, T. (2010). Purification and some properties of wild-type and N-terminal-truncated ethanolamine ammonia-lyase of Escherichia coli. *J Biochem*, 147:83–93.

Akopian, D., Shen, K., Zhang, X., and Shan, S.-o. (2013). Signal recognition particle: an essential protein-targeting machine. *Annu Rev Biochem*, 82:693–721.

Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. *Nature*, 426:570–574.

Anderson, N. (1955). Studies on isolated cell components. *Exp Cell Res*, 9(3):446–459.

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

Andresen, L. and Holmqvist, E. (2018). CLIP-Seq in Bacteria: Global Recognition Patterns of Bacterial RNA-Binding Proteins. *Methods Enzymol*, 612:127–145.

Aravind, L. and Koonin, E. V. (1998). The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem Sci*, 23:469–472.

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

Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H.-C., Hirai, A., Tsuzuki, K., Nakamura, S., Altaf-Ul-Amin, M., Oshima, T., Baba, T., Yamamoto, N., Kawamura, T., Ioka-Nakamichi, T., Kitagawa, M., Tomita, M., Kanaya, S., Wada, C., and Mori, H. (2006). Large-scale identification of protein-protein interaction of Escherichia coli K-12. *Genome Res*, 16:686–691.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25:25–29.

Atkinson, G. C., Tenson, T., and Hauryliuk, V. (2011). The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. *PLoS One*, 6:e23479.

Avery, O. T., MacLeod, C. M., and Maclyn, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *J Exp Med*, 79(2):137–158.

Babitzke, P., Lai, Y.-J., Renda, A. J., and Romeo, T. (2019). Posttranscription Initiation Control of Gene Expression Mediated by Bacterial RNA-Binding Proteins. *Annu Rev Microbiol*, 73:43–67.

Babu, M., Bundalovic-Torma, C., Calmettes, C., Phanse, S., Zhang, Q., Jiang, Y., Minic, Z., Kim, S., Mehla, J., Gagarinova, A., Rodionova, I., Kumar, A., Guo, H., Kagan, O., Pogoutse, O., Aoki, H., Deineko, V., Caufield, J. H., Holtzapple, E., Zhang, Z., Vastermark, A., Pandya, Y., Lai, C. C.-L., El Bakkouri, M., Hooda, Y., Shah, M., Burnside, D., Hooshyar, M., Vlasblom, J., Rajagopala, S. V., Golshani, A., Wuchty, S., F Greenblatt, J., Saier, M., Uetz, P., F Moraes, T., Parkinson, J., and Emili, A. (2018). Global landscape of cell envelope protein complexes in Escherichia coli. *Nat Biotechnol*, 36:103–112.

Bak, G., Lee, J., Suk, S., Kim, D., Young Lee, J., Kim, K.-S., Choi, B.-S., and Lee, Y. (2015). Identification of novel sRNAs involved in biofilm formation, motility, and fimbriae formation in Escherichia coli. *Sci Rep*, 5:15287.

Baker, S. C., Bauer, S. R., Beyer, R. P., Brenton, J. D., Bromley, B., Burrill, J., Causton, H., Conley, M. P., Elespuru, R., Fero, M., Foy, C., Fuscoe, J., Gao, X., Gerhold, D. L., Gilles, P., Goodsaid, F., Guo, X., Hackett, J., Hockett, R. D., Ikonomi, P.,

Irizarry, R. A., Kawasaki, E. S., Kaysser-Kranich, T., Kerr, K., Kiser, G., Koch, W. H., Lee, K. Y., Liu, C., Liu, Z. L., Lucas, A., Manohar, C. F., Miyada, G., Modrusan, Z., Parkes, H., Puri, R. K., Reid, L., Ryder, T. B., Salit, M., Samaha, R. R., Scherf, U., Sendera, T. J., Setterquist, R. A., Shi, L., Shippy, R., Soriano, J. V., Wagar, E. A., Warrington, J. A., Williams, M., Wilmer, F., Wilson, M., Wolber, P. K., Wu, X., Zadro, R., and Consortium, E. R. C. (2005). The External RNA Controls Consortium: a progress report. *Nat Methods*, 2:731–734.

Balasubramanian, D. and Vanderpool, C. K. (2013). New developments in post-transcriptional regulation of operons by small RNAs. *RNA Biol*, 10:337–341.

Bandyra, K. J., Bouvier, M., Carpousis, A. J., and Luisi, B. F. (2013). The social fabric of the RNA degradosome. *Biochim Biophys Acta*, 1829:514–522.

Bandyra, K. J., Sinha, D., Syrjanen, J., Luisi, B. F., and De Lay, N. R. (2016). The ribonuclease polynucleotide phosphorylase can interact with small regulatory RNAs in both protective and degradative modes. *RNA*, 22:360–372.

Bardwell, V. J. and Wickens, M. (1990). Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res*, 18:6587–6594.

Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-mediated general stress response in Escherichia coli. *Annu Rev Microbiol*, 65:189–213.

Bauriedl, S., Gerovac, M., Heidrich, N., Bischler, T., Barquist, L., Vogel, J., and Schoen, C. (2020). Functional characterization of the minimal ProQ protein of Neisseria meningitidis. *Nat Commun*.

Bechhofer, D. H. and Deutscher, M. P. (2019). Bacterial ribonucleases and their roles in RNA metabolism. *Crit Rev Biochem Mol Biol*, 54:242–300.

Berggård, T., Linse, S., and James, P. (2007). Methods for the detection and analysis of protein-protein interactions. *Proteomics*, 7:2833–2842.

Berghoff, B. A. and Wagner, E. G. H. (2017). RNA-based regulation in type I toxin-antitoxin systems and its implication for bacterial persistence. *Curr Genet*, 63:1011–1016.

Berndt, H., Harnisch, C., Rammelt, C., Stöhr, N., Zirkel, A., Dohm, J. C., Himmelbauer, H., Tavanez, J.-P., Hüttelmaier, S., and Wahle, E. (2012). Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA*, 18:958–972.

Bertrand, R. L. (2019). Lag Phase Is a Dynamic, Organized, Adaptive, and Evolvable Period That Prepares Bacteria for Cell Division. *J Bacteriol*, 201.

Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H., and Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, 452:846–850.

Bhandari, V. and Houry, W. A. (2015).   Substrate Interaction Networks of the Escherichia coli Chaperones: Trigger Factor, DnaK and GroEL. *Adv Exp Med Biol*, 883:271–294.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science*, 277:1453–1462.

Bobay, L.-M., Touchon, M., and Rocha, E. P. C. (2014). Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci U S A*, 111:12127–12132.

Boudry, P., Gracia, C., Monot, M., Caillet, J., Saujet, L., Hajnsdorf, E., Dupuy, B., Martin-Verstraete, I., and Soutourina, O. (2014).   Pleiotropic role of the RNA chaperone protein Hfq in the human pathogen Clostridium difficile. *J Bacteriol*, 196:3234–3248.

Brakke, M. K. (1951).   Density Gradient Centrifugation: A New Separation Technique1. *J Am Chem Soc*, 73(4):1847–1848.

Brakke, M. K. (1961). Density Gradient Centrifugation And its Application to Plant Viruses. In *Advances in Virus Research Volume 7*, pages 193–224. Elsevier.

Brantl, S. and Brückner, R. (2014).   Small regulatory RNAs from low-GC Gram-positive bacteria. *RNA Biol*, 11:443–456.

Braun, V., Mahren, S., and Ogierman, M. (2003).   Regulation of the FecI-type ECF sigma factor by transmembrane signalling. *Curr Opin Microbiol*, 6:173–180.

Brückner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009).   Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci*, 10:2763–2788.

Bremer, H. and Dennis, P. P. (2008).   Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, doi:10.1128/ecosal.5.2.3.

Bronesky, D., Wu, Z., Marzi, S., Walter, P., Geissmann, T., Moreau, K., Vandenesch, F., Caldelari, I., and Romby, P. (2016). Staphylococcus aureus RNAIII and Its Regulon Link Quorum Sensing, Stress Responses, Metabolic Adaptation, and Regulation of Virulence Gene Expression. *Annu Rev Microbiol*, 70:299–316.

Bättig, P., Hathaway, L. J., Hofer, S., and Mühlemann, K. (2006).   Serotype-specific invasiveness and colonization prevalence in Streptococcus pneumoniae correlate with the lag phase during in vitro growth. *Microbes Infect*, 8:2612–2617.

Burakovsky, D. E., Prokhorova, I. V., Sergiev, P. V., Milón, P., Sergeeva, O. V., Bogdanov, A. A., Rodnina, M. V., and Dontsova, O. A. (2012). Impact of methylations of m2G966/m5C967 in 16S rRNA on bacterial fitness and translation initiation. *Nucleic Acids Res*, 40:7885–7895.

Burenina, O. Y., Hoch, P. G., Damm, K., Salas, M., Zatsepin, T. S., Lechner, M., Oretskaya, T. S., Kubareva, E. A., and Hartmann, R. K. (2014). Mechanistic comparison of Bacillus subtilis 6S-1 and 6S-2 RNAs–commonalities and differences. *RNA*, 20:348–359.

Burkhardt, D. H., Rouskin, S., Zhang, Y., Li, G.-W., Weissman, J. S., and Gross, C. A. (2017). Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *eLife*, 6:e22037.

Burr, T., Mitchell, J., Kolb, A., Minchin, S., and Busby, S. (2000). DNA sequence elements located immediately upstream of the -10 hexamer in Escherichia coli promoters: a systematic study. *Nucleic Acids Res*, 28:1864–1870.

Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., and Emili, A. (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature*, 433:531–537.

Caillet, J., Gracia, C., Fontaine, F., and Hajnsdorf, E. (2014). Clostridium difficile Hfq can replace Escherichia coli Hfq for most of its function. *RNA*, 20:1567–1578.

Cameron, T. A. and De Lay, N. R. (2016). The Phosphorolytic Exoribonucleases Polynucleotide Phosphorylase and RNase PH Stabilize sRNAs and Facilitate Regulation of Their mRNA Targets. *J Bacteriol*, 198:3309–3317.

Cameron, T. A., Matz, L. M., and De Lay, N. R. (2018). Polynucleotide phosphorylase: Not merely an RNase but a pivotal post-transcriptional regulator. *PLoS Genet*, 14:e1007654.

Cameron, T. A., Matz, L. M., Sinha, D., and De Lay, N. R. (2019). Polynucleotide phosphorylase promotes the stability and function of Hfq-binding sRNAs by degrading target mRNA-derived fragments. *Nucleic Acids Res*, 47:8821–8837.

Carlson, M. L., Stacey, R. G., Young, J. W., Wason, I. S., Zhao, Z., Rattray, D. G., Scott, N., Kerr, C. H., Babu, M., Foster, L. J., and Duong, F. V. H. (2019). Profiling the E. coli membrane interactome captured in peptidisc libraries. *eLife*, 8:e46615.

Cassone, M., Gagne, A. L., Spruce, L. A., Seeholzer, S. H., and Sebert, M. E. (2012). The HtrA protease from Streptococcus pneumoniae digests both denatured proteins and the competence-stimulating peptide. *J Biol Chem*, 287:38449–38459.

Caudron-Herger, M., Rusin, S. F., Adamo, M. E., Seiler, J., Schmid, V. K., Barreau, E., Kettenbach, A. N., and Diederichs, S. (2019). R-DeeP: Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation. *Mol Cell*, 75:184–199.e10.

Caudron-Herger, M., Wassmer, E., Nasa, I., Schultz, A.-S., Seiler, J., Kettenbach, A. N., and Diederichs, S. (2020). Identification, quantification and bioinformatic analysis of RNA-dependent proteins by RNase treatment and density gradient ultracentrifugation using R-DeeP. *Nat Protoc*, doi:10.1038/s41596-019-0261-4.

Cavanagh, A. T., Sperger, J. M., and Wassarman, K. M. (2012). Regulation of 6S RNA by pRNA synthesis is required for efficient recovery from stationary phase in E. coli and B. subtilis. *Nucleic Acids Res*, 40:2234–2246.

Cech, T. R. and Steitz, J. A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, 157:77–94.

Chalker, A. F., Ingraham, K. A., Lunsford, R. D., Bryant, A. P., Bryant, J., Wallis, N. G., Broskey, J. P., Pearson, S. C., and Holmes, D. J. (2000). The bacA gene, which determines bacitracin susceptibility in Streptococcus pneumoniae and Staphylococcus aureus, is also required for virulence. *Microbiology*, 146 (Pt 7):1547–1553.

Champ, S., Puvirajesinghe, T. M., Perrody, E., Menouni, R., Genevaux, P., and Ansaldi, M. (2011). Chaperone-assisted excisive recombination, a solitary role for DnaJ (Hsp40) chaperone in lysogeny escape. *J Biol Chem*, 286:38876–38885.

Chen, C.-S., Korobkova, E., Chen, H., Zhu, J., Jian, X., Tao, S.-C., He, C., and Zhu, H. (2008). A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. *Nat Methods*, 5:69–74.

Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994). Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res*, 22:4953–4957.

Chen, S., Lesnik, E. A., Hall, T. A., Sampath, R., Griffey, R. H., Ecker, D. J., and Blyn, L. B. (2002). A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. *Biosystems*, 65:157–177.

Chen, S. S. and Williamson, J. R. (2013). Characterization of the ribosome biogenesis landscape in E. coli using quantitative mass spectrometry. *J Mol Biol*, 425:767–779.

Cherepanov, P. P. and Wackernagel, W. (1995). Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*, 158:9–14.

Chihara, K., Bischler, T., Barquist, L., Monzon, V. A., Noda, N., Vogel, J., and Tsuneda, S. (2019). Conditional Hfq Association with Small Noncoding RNAs in Pseudomonas aeruginosa Revealed through Comparative UV Cross-Linking Immunoprecipitation Followed by High-Throughput Sequencing. *mSystems*, 4.

Cho, K. H. (2017). The Structure and Function of the Gram-Positive Bacterial RNA Degradosome. *Front Microbiol*, 8:154.

Choi, J. S., Kim, W., Suk, S., Park, H., Bak, G., Yoon, J., and Lee, Y. (2018). The small RNA, SdsR, acts as a novel type of toxin in Escherichia coli. *RNA Biol*, 15:1319–1335.

Claude, A. (1946a). Fractionation of Mammalian Liver Cells by Differential Centrifugation: I. Problems, Methods, and Preparation of Extract. *J Exp Med*, 84(1):51–59.

Claude, A. (1946b). Fractionation of Mammalian Liver Cells by Differential Centrifugation: II. Experimental Procedures and Results. *J Exp Med*, 84(1):61–89.

Conway, T., Creecy, J. P., Maddox, S. M., Grissom, J. E., Conkle, T. L., Shadid, T. M., Teramoto, J., San Miguel, P., Shimada, T., Ishihama, A., Mori, H., and Wanner, B. L. (2014). Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *mBio*, 5:e01442–e01414.

Corcoran, C. P., Rieder, R., Podkaminski, D., Hofmann, B., and Vogel, J. (2012). Use of aptamer tagging to identify in vivo protein binding partners of small regulatory RNAs. *Methods Mol Biol*, 905:177–200.

Cossa, G., Roeschert, I., Prinz, F., Baluapuri, A., Silveira Vidal, R., Schülein-Völk, C., Chang, Y.-C., Ade, C. P., Mastrobuoni, G., Girard, C., Wortmann, L., Walz, S., Lührmann, R., Kempa, S., Kuster, B., Wolf, E., Mumberg, D., and Eilers, M. (2020). Localized Inhibition of Protein Phosphatase 1 by NUAK1 Promotes Spliceosome Activity and Reveals a MYC-Sensitive Feedback Control of Transcription. *Mol Cell*, 10.1016/j.molcel.2020.01.008.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*, 13:2513–2526.

Cox, J. and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*, 26:1367–1372.

Crick, F. H. (1958). On protein synthesis. *Symp Soc Exp Biol*, 12:138–163.

Crick, F. H. (1970). Central dogma of molecular biology. *Nature*, 227:561–563.

Cronan, J. E. and Waldrop, G. L. (2002). Multi-subunit acetyl-CoA carboxylases. *Prog Lipid Res*, 41:407–435.

Crozier, T. W. M., Tinti, M., Larance, M., Lamond, A. I., and Ferguson, M. A. J. (2017). Prediction of Protein Complexes in Trypanosoma brucei by Protein Correlation Profiling Mass Spectrometry and Machine Learning. *Mol Cell Proteomics*, 16:2254–2267.

Dambach, M., Irnov, I., and Winkler, W. C. (2013). Association of RNAs with Bacillus subtilis Hfq. *PLoS One*, 8:e55156.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25:1974–1975.

Datsenko, K. A. and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A*, 97:6640–6645.

Davis, J. H. and Williamson, J. R. (2017). Structure and dynamics of bacterial ribosome biogenesis. *Philos Trans R Soc Lond B Biol Sci*, 372:20160181.

De Lay, N. and Gottesman, S. (2011). Role of polynucleotide phosphorylase in sRNA function in Escherichia coli. *RNA*, 17:1172–1189.

Desgranges, E., Marzi, S., Moreau, K., Romby, P., and Caldelari, I. (2019). Noncoding RNA. *Microbiol Spectr*, 7:GPP3–0038–2018.

Diéguez-Casal, E., Freixeiro, P., Costoya, L., Criado, M. T., Ferreirós, C., and Sánchez, S. (2014). High resolution clear native electrophoresis is a good alternative to blue native electrophoresis for the characterization of the Escherichia coli membrane complexes. *J Microbiol Methods*, 102:45–54.

Doerfel, L. K., Wohlgemuth, I., Kothe, C., Peske, F., Urlaub, H., and Rodnina, M. V. (2013). EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science*, 339:85–88.

Durand, S. and Condon, C. (2018). RNases and helicases in gram-positive bacteria. *Microbiol Spectr*, 6(2):RWR–0003–2017.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., and Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res*, 47:D427–D432.

Erickson, H. P. (2009). Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online*, 11:32–51.

Fang, M., Zeisberg, W.-M., Condon, C., Ogryzko, V., Danchin, A., and Mechold, U. (2009). Degradation of nanoRNA is performed by multiple redundant RNases in Bacillus subtilis. *Nucleic Acids Res*, 37:5114–5125.

Feklístov, A., Sharon, B. D., Darst, S. A., and Gross, C. A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol*, 68:357–376.

Feng, L., Rutherford, S. T., Papenfort, K., Bagert, J. D., van Kessel, J. C., Tirrell, D. A., Wingreen, N. S., and Bassler, B. L. (2015). A qrr noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell*, 160:228–240.

Fijalkowska, I. J., Schaaper, R. M., and Jonczyk, P. (2012). DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair. *FEMS Microbiol Rev*, 36:1105–1121.

Flynn, R. L. and Zou, L. (2010). Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. *Crit Rev Biochem Mol Biol*, 45:266–275.

Foster, L. J., de Hoog, C. L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V. K., and Mann, M. (2006). A mammalian organelle map by protein correlation profiling. *Cell*, 125:187–199.

Fröhlich, K. S., Papenfort, K., Fekete, A., and Vogel, J. (2013). A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J*, 32:2963–2979.

Fridman, O., Goldberg, A., Ronin, I., Shoresh, N., and Balaban, N. Q. (2014). Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature*, 513:418–421.

Förstner, K. U., Vogel, J., and Sharma, C. M. (2014). READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics*, 30:3421–3423.

Gazestani, V. H., Nikpour, N., Mehta, V., Najafabadi, H. S., Moshiri, H., Jardim, A., and Salavati, R. (2016). A Protein Complex Map of Trypanosoma brucei. *PLoS Negl Trop Dis*, 10:e0004533.

Gekko, K. and Timasheff, S. N. (1981a). Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. *Biochemistry*, 20:4667–4676.

Gekko, K. and Timasheff, S. N. (1981b). Thermodynamic and kinetic examination of protein stabilization by glycerol. *Biochemistry*, 20:4677–4686.

Geno, K. A., Gilbert, G. L., Song, J. Y., Skovsted, I. C., Klugman, K. P., Jones, C., Konradsen, H. B., and Nahm, M. H. (2015). Pneumococcal Capsules and Their Types: Past, Present, and Future. *Clin Microbiol Rev*, 28:871–899.

Gerdes, K. and Wagner, E. G. H. (2007). RNA antitoxins. *Curr Opin Microbiol*, 10:117–124.

Gómez-Mejia, A., Gámez, G., and Hammerschmidt, S. (2018). Streptococcus pneumoniae two-component regulatory systems: The interplay of the pneumococcus with its environment. *Int J Med Microbiol*, 308:722–737.

Gorka, M., Swart, C., Siemiatkowska, B., Martínez-Jaime, S., Skirycz, A., Streb, S., and Graf, A. (2019). Protein Complex Identification and quantitative complexome by CN-PAGE. *Sci Rep*, 9:11523.

Gorski, S. A., Vogel, J., and Doudna, J. A. (2017). RNA-based recognition and targeting: sowing the seeds of specificity. *Nat Rev Mol Cell Biol*, 18:215–228.

Griffith, F. (1928). The Significance of Pneumococcal Types. *J Hyg (Lond)*, 27(2):113–159.

Gualerzi, C. O. and Pon, C. L. (2015). Initiation of mRNA translation in bacteria: structural and dynamic aspects. *Cell Mol Life Sci*, 72:4341–4367.

Halfmann, A., Kovács, M., Hakenbeck, R., and Brückner, R. (2007). Identification of the genes directly controlled by the response regulator CiaR in Streptococcus pneumoniae: five out of 15 promoters drive expression of small non-coding RNAs. *Mol Microbiol*, 66:110–126.

Halpern, D., Gruss, A., Claverys, J.-P., and Karoui, M. E. (2004). rexAB mutants in Streptococcus pneumoniae. *Microbiology*, 150:2409–2414.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., Huang, D., Xu, Y., Huang, W., Jiang, M., Jiang, X., Mao, J., Chen, Y., Lu, C., Xie, J., Fang, Q., Wang, Y., Yue, R., Li, T., Huang, H., Orkin, S. H., Yuan, G.-C., Chen, M., and Guo, G. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*, 172:1091–1107.e17.

Harms, A., Brodersen, D. E., Mitarai, N., and Gerdes, K. (2018). Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol Cell*, 70:768–784.

Hartz, D., McPheeters, D. S., Traut, R., and Gold, L. (1988). Extension inhibition analysis of translation initiation complexes. *Methods Enzymol*, 164:419–425.

Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., Wang, P. I., Boutz, D. R., Fong, V., Phanse, S., Babu, M., Craig, S. A., Hu, P., Wan, C., Vlasblom, J., Dar, V.-u.-N., Bezginov, A., Clark, G. W., Wu, G. C., Wodak, S. J., Tillier, E. R. M., Paccanaro, A., Marcotte, E. M., and Emili, A. (2012). A census of human soluble protein complexes. *Cell*, 150:1068–1081.

Helm, M. and Motorin, Y. (2017). Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat Rev Genet*, 18:275–291.

Henriques-Normark, B. and Tuomanen, E. I. (2013). The pneumococcus: epidemiology, microbiology, and pathogenesis. *Cold Spring Harb Perspect Med*, 3:a010215.

Hensley, M. P., Gunasekera, T. S., Easton, J. A., Sigdel, T. K., Sugarbaker, S. A., Klingbeil, L., Breece, R. M., Tierney, D. L., and Crowder, M. W. (2012). Characterization of Zn(II)-responsive ribosomal proteins YkgM and L31 in E. coli. *J Inorg Biochem*, 111:164–172.

Hiller, N. L. and Sá-Leão, R. (2018). Puzzling Over the Pneumococcal Pangenome. *Front Microbiol*, 9:2580.

Hämmerle, H., Amman, F., Večerek, B., Stülke, J., Hofacker, I., and Bläsi, U. (2014). Impact of Hfq on the Bacillus subtilis transcriptome. *PLoS One*, 9:e98661.

Hoffmann, A., Bukau, B., and Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. *Biochim Biophys Acta*, 1803:650–661.

Hoffmann, S., Otto, C., Doose, G., Tanzer, A., Langenberger, D., Christ, S., Kunz, M., Holdt, L. M., Teupser, D., Hackermüller, J., and Stadler, P. F. (2014). A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol*, 15:R34.

Hogeboom, G. H., Schneider, W. C., and Pallade, G. E. (1948). Cytochemical studies of mammalian tissues; isolation of intact mitochondria from rat liver; some biochemical properties of mitochondria and submicroscopic particulate material. *J Biol Chem*, 172:619–635.

Holmqvist, E., Li, L., Bischler, T., Barquist, L., and Vogel, J. (2018). Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3' Ends. *Mol Cell*, 70:971–982.e6.

Holmqvist, E. and Vogel, J. (2018). RNA-binding proteins in bacteria. *Nat Rev Microbiol*, 16:601–615.

Hong, P., Koza, S., and Bouvier, E. S. P. (2012). Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol*, 35:2923–2950.

Hör, J., Garriss, G., Di Giorgio, S., Hack, L.-M., Vanselow, J. T., Förstner, K. U., Schlosser, A., Henriques-Normark, B., and Vogel, J. (2020a). Grad-seq in a Gram-positive bacterium reveals exonucleolytic sRNA activation in competence control. *EMBO J*, doi:10.15252/embj.2019103852.

Hör, J., Gorski, S. A., and Vogel, J. (2018). Bacterial RNA Biology on a Genome Scale. *Mol Cell*, 70:785–799.

Hör, J., Matera, G., Vogel, J., Gottesman, S., and Storz, G. (2020b). Trans-Acting Small RNAs and Their Effects on Gene Expression in Escherichia coli and Salmonella enterica. *EcoSal Plus*, doi:10.1128/ecosalplus.ESP-0030-2019.

Hör, J. and Vogel, J. (2017). Global snapshots of bacterial RNA networks. *EMBO J*, 36:245–247.

Hu, P., Janga, S. C., Babu, M., Díaz-Mejía, J. J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S., Wong, P., Chandran, S., Christopoulos, C., Nazarians-Armavil, A., Nasseri, N. K., Musso, G., Ali, M., Nazemof, N., Eroukova, V., Golshani, A.,

Paccanaro, A., Greenblatt, J. F., Moreno-Hagelsieb, G., and Emili, A. (2009). Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins. *PLoS Biol*, 7:e96.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., von Mering, C., and Bork, P. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*, 44:D286–D293.

Häuser, R., Pech, M., Kijek, J., Yamamoto, H., Titz, B., Naeve, F., Tovchigrechko, A., Yamamoto, K., Szaflarski, W., Takeuchi, N., Stellberger, T., Diefenbacher, M. E., Nierhaus, K. H., and Uetz, P. (2012). RsfA (YbeB) proteins are conserved ribosomal silencing factors. *PLos Genet*, 8:e1002815.

Inouye, M. and Delihas, N. (1988). Small RNAs in the prokaryotes: a growing list of diverse roles. *Cell*, 53:5–7.

Jiang, L., Schaffitzel, C., Bingel-Erlenmeyer, R., Ban, N., Korber, P., Koning, R. I., de Geus, D. C., Plaisier, J. R., and Abrahams, J. P. (2009). Recycling of aborted ribosomal 50S subunit-nascent chain-tRNA complexes by the heat shock protein Hsp15. *J Mol Biol*, 386:1357–1367.

Jiang, M., Datta, K., Walker, A., Strahler, J., Bagamasbad, P., Andrews, P. C., and Maddock, J. R. (2006). The Escherichia coli GTPase CgtAE is involved in late steps of large ribosome assembly. *J Bacteriol*, 188:6757–6770.

Jiang, M., Sullivan, S. M., Walker, A. K., Strahler, J. R., Andrews, P. C., and Maddock, J. R. (2007). Identification of novel Escherichia coli ribosome-associated proteins using isobaric tags and multidimensional protein identification techniques. *J Bacteriol*, 189:3434–3444.

Jomaa, A., Jain, N., Davis, J. H., Williamson, J. R., Britton, R. A., and Ortega, J. (2014). Functional domains of the 50S subunit mature late in the assembly process. *Nucleic Acids Res*, 42:3419–3435.

Jones, A. L., Needham, R. H. V., and Rubens, C. E. (2003). The Delta subunit of RNA polymerase is required for virulence of Streptococcus agalactiae. *Infect Immun*, 71:4011–4017.

Junge, W. and Nelson, N. (2015). ATP synthase. *Annu Rev Biochem*, 84:631–657.

Kajitani, M., Kato, A., Wada, A., Inokuchi, Y., and Ishihama, A. (1994). Regulation of the Escherichia coli hfq gene encoding the host factor for phage Q beta. *J Bacteriol*, 176:531–534.

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., and Subhraveti, P. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings Bioinf*, 20:1085–1093.

Karzai, A. W., Roche, E. D., and Sauer, R. T. (2000). The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol*, 7:449–455.

Kavita, K., de Mets, F., and Gottesman, S. (2018). New aspects of RNA-based regulation by Hfq and its partner sRNAs. *Curr Opin Microbiol*, 42:53–61.

Keiler, K. C. (2015). Mechanisms of ribosome rescue in bacteria. *Nat Rev Microbiol*, 13:285–297.

Kersters, K., De Vos, P., Gillis, M., Swings, J., Vandamme, P., and Stackebrandt, E. (2006). Introduction to the Proteobacteria. In Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., editors, *The Prokaryotes*, volume 5 : Proteobacteria : alpha and beta subclasses, pages 3–37. Springer.

Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., Latendresse, M., Muñiz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velázquez-Ramírez, D. A., Weaver, D., Collado-Vides, J., Paulsen, I., and Karp, P. D. (2017). The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Res*, 45:D543–D550.

Kilian, M. and Tettelin, H. (2019). Identification of Virulence-Associated Properties by Comparative Genome Analysis of Streptococcus pneumoniae, S. pseudopneumoniae, S. mitis, Three S. oralis Subspecies, and S. infantis. *mBio*, 10:e01985–19.

Kim, B., Jeong, K., and Kim, V. N. (2017). Genome-wide Mapping of DROSHA Cleavage Sites on Primary MicroRNAs and Noncanonical Substrates. *Mol Cell*, 66:258–269.e5.

Kim, B. and Kim, V. N. (2019). fCLIP-seq for transcriptomic footprinting of dsRNA-binding proteins: Lessons from DROSHA. *Methods*, 152:3–11.

Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2011). Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods*, 9:72–74.

Koegl, M. and Uetz, P. (2007). Improving yeast two-hybrid screening systems. *Brief Funct Genomic Proteomic*, 6:302–312.

Korber, P., Stahl, J. M., Nierhaus, K. H., and Bardwell, J. C. (2000). Hsp15: a ribosome-associated heat shock protein. *EMBO J*, 19:741–748.

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31:147–157.

Kudla, G., Granneman, S., Hahn, D., Beggs, J. D., and Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc Natl Acad Sci U S A*, 108:10010–10015.

Kumar, R., Shah, P., Swiatlo, E., Burgess, S. C., Lawrence, M. L., and Nanduri, B. (2010). Identification of novel non-coding small RNAs from Streptococcus pneumoniae TIGR4 using high-resolution genome tiling arrays. *BMC Genomics*, 11:350.

Kwan, J. H. M. and Emili, A. (2016). Simple and Effective Affinity Purification Procedures for Mass Spectrometry-Based Identification of Protein-Protein Interactions in Cell Signaling Pathways. *Methods Mol Biol*, 1394:181–187.

Lacks, S. and Hotchkiss, R. D. (1960). A study of the genetic material determining an enzyme in Pneumococcus. *Biochim Biophys Acta*, 39:508–518.

Lalaouna, D., Carrier, M.-C., Semsey, S., Brouard, J.-S., Wang, J., Wade, J. T., and Massé, 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.

Lasserre, J.-P., Beyne, E., Pyndiah, S., Lapaillerie, D., Claverol, S., and Bonneu, M. (2006). A complexomic study of Escherichia coli using two-dimensional blue native/SDS polyacrylamide gel electrophoresis. *Electrophoresis*, 27:3306–3321.

Laux, A., Sexauer, A., Sivaselvarajah, D., Kaysen, A., and Brückner, R. (2015). Control of competence by related non-coding csRNAs in Streptococcus pneumoniae R6. *Front Genet*, 6:246.

Lécrivain, A.-L., Le Rhun, A., Renault, T. T., Ahmed-Begrich, R., Hahnke, K., and Charpentier, E. (2018). In vivo 3'-to-5' exoribonuclease targetomes of Streptococcus pyogenes. *Proc Natl Acad Sci U S A*, 115:11814–11819.

Lease, R. A., Cusick, M. E., and Belfort, M. (1998). Riboregulation in Escherichia coli: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc Natl Acad Sci U S A*, 95:12456–12461.

Lee, F. C. Y. and Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol Cell*, 69:354–369.

Lee, J. C. and Timasheff, S. N. (1981). The stabilization of proteins by sucrose. *J Biol Chem*, 256:7193–7201.

Lesnyak, D. V., Osipiuk, J., Skarina, T., Sergiev, P. V., Bogdanov, A. A., Edwards, A., Savchenko, A., Joachimiak, A., and Dontsova, O. A. (2007). Methyltransferase that modifies guanine 966 of the 16 S rRNA: functional identification and tertiary structure. *J Biol Chem*, 282:5880–5887.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157:624–635.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25:402–408.

Lloyd, C. R., Park, S., Fei, J., and Vanderpool, C. K. (2017). The Small Protein SgrT Controls Transport Activity of the Glucose-Specific Phosphotransferase System. *J Bacteriol*, 199:e00869–16.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Trans Inf Theory*, 28(2):129–137.

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.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15:550.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*, 47:W636–W641.

Małecka, E. M., Stróżecka, J., Sobańska, D., and Olejniczak, M. (2015). Structure of bacterial regulatory RNAs determines their performance in competition for the chaperone protein Hfq. *Biochemistry*, 54:1157–1170.

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., and 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.

Mann, B., van Opijnen, T., Wang, J., Obert, C., Wang, Y.-D., Carter, R., McGoldrick, D. J., Ridout, G., Camilli, A., Tuomanen, E. I., and Rosch, J. W. (2012). Control of virulence by small RNAs in Streptococcus pneumoniae. *PLoS Pathog*, 8:e1002788.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10.

Marx, P., Nuhn, M., Kovács, M., Hakenbeck, R., and Brückner, R. (2010). Identification of genes for small non-coding RNAs that belong to the regulon of the two-component regulatory system CiaRH in Streptococcus. *BMC Genomics*, 11:661.

Mathew, R. and Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol*, 14:450–455.

Mazumder, A. and Kapanidis, A. N. (2019). Recent Advances in Understanding Sigma70-Dependent Transcription Initiation Mechanisms. *J Mol Biol*, 431:3947–3959.

McIntosh, B. K., Renfro, D. P., Knapp, G. S., Lairikyengbam, C. R., Liles, N. M., Niu, L., Supak, A. M., Venkatraman, A., Zweifel, A. E., Siegele, D. A., and Hu, J. C. (2012). EcoliWiki: a wiki-based community resource for Escherichia coli. *Nucleic Acids Res*, 40:D1270–D1277.

Meades, G., Benson, B. K., Grove, A., and Waldrop, G. L. (2010). A tale of two functions: enzymatic activity and translational repression by carboxyltransferase. *Nucleic Acids Res*, 38:1217–1227.

Mehta, P., Woo, P., Venkataraman, K., and Karzai, A. W. (2012). Ribosome purification approaches for studying interactions of regulatory proteins and RNAs with the ribosome. *Methods Mol Biol*, 905:273–289.

Melamed, S., Adams, P. P., Zhang, A., Zhang, H., and Storz, G. (2020). RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles. *Mol Cell*, 77:411–425.e7.

Melamed, S., Faigenbaum-Romm, R., Peer, A., Reiss, N., Shechter, O., Bar, A., Altuvia, Y., Argaman, L., and Margalit, H. (2018). Mapping the small RNA interactome in bacteria using RIL-seq. *Nat Protoc*, 13:1–33.

Meselson, M. and Stahl, F. W. (1958). The Replication of DNA in Escherichia coli. *Proc Natl Acad Sci U S A*, 44:671–682.

Meselson, M., Stahl, F. W., and Vinograd, J. (1957). Equilibrium Sedimentation of Macromolecules in Density Gradients. *Proc Natl Acad Sci U S A*, 43:581–588.

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H.-Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D. A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A. P., Paysan-Lafosse, T., Pesseat, S., Potter, S. C., Qureshi, M. A., Rawlings, N. D., Redaschi, N., Richardson, L. J., Rivoire, C., Salazar, G. A., Sangrador-Vegas, A., Sigrist, C. J. A., Sillitoe, I., Sutton, G. G., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Yong, S.-Y., and Finn, R. D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*, 47:D351–D360.

Mitchell, J. E., Zheng, D., Busby, S. J. W., and Minchin, S. D. (2003). Identification and analysis of 'extended -10' promoters in Escherichia coli. *Nucleic Acids Res*, 31:4689–4695.

Miyakoshi, M., Chao, Y., and Vogel, J. (2015). Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Curr Opin Microbiol*, 24:132–139.

Müller, C. S., Bildl, W., Haupt, A., Ellenrieder, L., Becker, T., Hunte, C., Fakler, B., and Schulte, U. (2016). Cryo-slicing Blue Native-Mass Spectrometry (csBN-MS), a Novel Technology for High Resolution Complexome Profiling. *Mol Cell Proteomics*, 15:669–681.

Müller, C. S., Bildl, W., Klugbauer, N., Haupt, A., Fakler, B., and Schulte, U. (2019). High-Resolution Complexome Profiling by Cryoslicing BN-MS Analysis. *J Vis Exp*, e60096.

Møller, T., Franch, T., Udesen, C., Gerdes, K., and Valentin-Hansen, P. (2002). Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. *Genes Dev*, 16:1696–1706.

Mohanty, B. K. and Kushner, S. R. (2018). Enzymes Involved in Posttranscriptional RNA Metabolism in Gram-Negative Bacteria. *Microbiol Spectr*, 6:RWR–0011–2017.

Mondragón, A. (2013). Structural studies of RNase P. *Annu Rev Biophys*, 42:537–557.

Moscoso, M. and Claverys, J.-P. (2004). Release of DNA into the medium by competent Streptococcus pneumoniae: kinetics, mechanism and stability of the liberated DNA. *Mol Microbiol*, 54:783–794.

Muschiol, S., Aschtgen, M.-S., Nannapaneni, P., and Henriques-Normark, B. (2019). Gram-Positive Type IV Pili and Competence. *Microbiol Spectr*, 7:PSIB–0011–2018.

Muschiol, S., Erlendsson, S., Aschtgen, M.-S., Oliveira, V., Schmieder, P., de Lichtenberg, C., Teilum, K., Boesen, T., Akbey, U., and Henriques-Normark, B. (2017). Structure of the competence pilus major pilin ComGC in Streptococcus pneumoniae. *J Biol Chem*, 292:14134–14146.

Nielsen, J. S., Lei, L. K., Ebersbach, T., Olsen, A. S., Klitgaard, J. K., Valentin-Hansen, P., and Kallipolitis, B. H. (2010). Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in Listeria monocytogenes. *Nucleic Acids Res*, 38:907–919.

Nikolay, R., van den Bruck, D., Achenbach, J., and Nierhaus, K. H. (2015). *Ribosomal Proteins: Role in Ribosomal Functions*, pages 1–12. American Cancer Society.

Noguchi, Y. and Katayama, T. (2016). The Escherichia coli Cryptic Prophage Protein YfdR Binds to DnaA and Initiation of Chromosomal Replication Is Inhibited by Overexpression of the Gene Cluster yfdQ-yfdR-yfdS-yfdT. *Front Microbiol*, 7:239.

O'Brien, K. L., Wolfson, L. J., Watt, J. P., Henkle, E., Deloria-Knoll, M., McCall, N., Lee, E., Mulholland, K., Levine, O. S., Cherian, T., Hib, and of Disease Study Team, P. G. B. (2009). Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *Lancet*, 374:893–902.

Omotajo, D., Tate, T., Cho, H., and Choudhary, M. (2015). Distribution and diversity of ribosome binding sites in prokaryotic genomes. *BMC Genomics*, 16:604.

Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wriggers, W., and Darst, S. A. (2003). Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell*, 114:335–345.

Oussenko, I. A., Abe, T., Ujiie, H., Muto, A., and Bechhofer, D. H. (2005). Participation of 3′-to-5′ exoribonucleases in the turnover of Bacillus subtilis mRNA. *J Bacteriol*, 187:2758–2767.

Oussenko, I. A. and Bechhofer, D. H. (2000). The yvaJ gene of Bacillus subtilis encodes a 3′-to-5′ exoribonuclease and is not essential in a strain lacking polynucleotide phosphorylase. *J Bacteriol*, 182:2639–2642.

Oussenko, I. A., Sanchez, R., and Bechhofer, D. H. (2002). Bacillus subtilis YhaM, a member of a new family of 3′-to-5′ exonucleases in gram-positive bacteria. *J Bacteriol*, 184:6250–6259.

Pan, J.-Y., Li, H., Ma, Y., Chen, P., Zhao, P., Wang, S.-Y., and Peng, X.-X. (2010). Complexome of Escherichia coli envelope proteins under normal physiological conditions. *J Proteome Res*, 9:3730–3740.

Pan, J.-Y., Wu, H., Liu, X., Li, P.-P., Li, H., Wang, S.-Y., and Peng, X.-X. (2011). Complexome of Escherichia coli cytosolic proteins under normal native conditions. *Mol Biosyst*, 7:2651–2663.

Panis, G., Duverger, Y., Champ, S., and Ansaldi, M. (2010a). Protein binding sites involved in the assembly of the KplE1 prophage intasome. *Virology*, 404:41–50.

Panis, G., Duverger, Y., Courvoisier-Dezord, E., Champ, S., Talla, E., and Ansaldi, M. (2010b). Tight regulation of the intS gene of the KplE1 prophage: a new paradigm for integrase gene regulation. *PLos Genet*, 6:e1001149.

Panis, G., Franche, N., Méjean, V., and Ansaldi, M. (2012). Insights into the functions of a prophage recombination directionality factor. *Viruses*, 4:2417–2431.

Panis, G., Méjean, V., and Ansaldi, M. (2007). Control and regulation of KplE1 prophage site-specific recombination: a new recombination module analyzed. *J Biol Chem*, 282:21798–21809.

Papenfort, K., Sun, Y., Miyakoshi, M., Vanderpool, C. K., and Vogel, J. (2013). Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell*, 153:426–437.

Papenfort, K. and Vanderpool, C. K. (2015). Target activation by regulatory RNAs in bacteria. *FEMS Microbiol Rev*, 39:362–378.

Pedersen, K. and Gerdes, K. (1999). Multiple hok genes on the chromosome of Escherichia coli. *Mol Microbiol*, 32:1090–1102.

Peterson, S. N., Sung, C. K., Cline, R., Desai, B. V., Snesrud, E. C., Luo, P., Walling, J., Li, H., Mintz, M., Tsegaye, G., Burr, P. C., Do, Y., Ahn, S., Gilbert, J., Fleischmann, R. D., and Morrison, D. A. (2004). Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. *Mol Microbiol*, 51:1051–1070.

Pircher, A., Gebetsberger, J., and Polacek, N. (2014). Ribosome-associated ncRNAs: an emerging class of translation regulators. *RNA Biol*, 11:1335–1339.

Pokkunuri, I. and Champney, W. S. (2007). Characteristics of a 50S ribosomal subunit precursor particle as a substrate for ermE methyltransferase activity and erythromycin binding in Staphylococcus aureus. *RNA Biol*, 4:147–153.

Potts, A. H., Vakulskas, C. A., Pannuri, A., Yakhnin, H., Babitzke, P., and Romeo, T. (2017). Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics. *Nat Commun*, 8:1596.

Pérez-Reytor, D., Plaza, N., Espejo, R. T., Navarrete, P., Bastías, R., and Garcia, K. (2016). Role of non-coding regulatory rna in the virulence of human pathogenic vibrios. *Front Microbiol*, 7:2160.

Puvirajesinghe, T. M., Elantak, L., Lignon, S., Franche, N., Ilbert, M., and Ansaldi, M. (2012). DnaJ (Hsp40 protein) binding to folded substrate impacts KplE1 prophage excision efficiency. *J Biol Chem*, 287:14169–14177.

Queiroz, R. M. L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R. F., Dezi, V., Thomas, G. H., Willis, A. E., and Lilley, K. S. (2019). Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat Biotechnol*, 37:169–178.

Quereda, J. J. and Cossart, P. (2017). Regulating Bacterial Virulence with RNA. *Annu Rev Microbiol*, 71:263–280.

Rabatinová, A., Šanderová, H., Jirát Matějčková, J., Korelusová, J., Sojka, L., Barvík, I., Papoušková, V., Sklenár, V., Žídek, L., and Krásný, L. (2013). The delta subunit of RNA polymerase is required for rapid changes in gene expression and competitive fitness of the cell. *J Bacteriol*, 195:2603–2611.

Rajagopala, S. V., Sikorski, P., Caufield, J. H., Tovchigrechko, A., and Uetz, P. (2012). Studying protein complexes by the yeast two-hybrid system. *Methods*, 58:392–399.

Rajagopala, S. V., Sikorski, P., Kumar, A., Mosca, R., Vlasblom, J., Arnold, R., Franca-Koh, J., Pakala, S. B., Phanse, S., Ceol, A., Häuser, R., Siszler, G., Wuchty, S., Emili,

A., Babu, M., Aloy, P., Pieper, R., and Uetz, P. (2014). The binary protein-protein interaction landscape of Escherichia coli. *Nat Biotechnol*, 32:285–290.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem*, 75:663–670.

Rederstorff, M., Bernhart, S. H., Tanzer, A., Zywicki, M., Perfler, K., Lukasser, M., Hofacker, I. L., and Hüttenhofer, A. (2010). RNPomics: defining the ncRNA transcriptome by cDNA library generation from ribonucleo-protein particles. *Nucleic Acids Res*, 38:e113.

Rederstorff, M. and Hüttenhofer, A. (2011). cDNA library generation from ribonucleoprotein particles. *Nat Protoc*, 6:166–174.

Redko, Y. and Condon, C. (2010). Maturation of 23S rRNA in Bacillus subtilis in the absence of Mini-III. *J Bacteriol*, 192:356–359.

Ricci, E. P., Kucukural, A., Cenik, C., Mercier, B. C., Singh, G., Heyer, E. E., Ashar-Patel, A., Peng, L., and Moore, M. J. (2014). Staufen1 senses overall transcript secondary structure to regulate translation. *Nat Struct Mol Biol*, 21:26–35.

Risso, D., Ngai, J., Speed, T. P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol*, 32:896–902.

Rizzatti, G., Lopetuso, L. R., Gibiino, G., Binda, C., and Gasbarrini, A. (2017). Proteobacteria: A Common Factor in Human Diseases. *Biomed Res Int*, 2017:9351507.

Rochat, T., Delumeau, O., Figueroa-Bossi, N., Noirot, P., Bossi, L., Dervyn, E., and Bouloc, P. (2015). Tracking the Elusive Function of Bacillus subtilis Hfq. *PLoS One*, 10:e0124977.

Romeo, T. and Babitzke, P. (2018). Global Regulation by CsrA and Its RNA Antagonists. *Microbiol Spectr*, 6:RWR–0009–2017.

Ron, E. Z., Kohler, R. E., and Davis, B. D. (1968). Magnesium ion dependence of free and polysomal ribosomes from Escherichia coli. *J Mol Biol*, 36:83–89.

Rugen, N., Straube, H., Franken, L. E., Braun, H.-P., and Eubel, H. (2019). Complexome profiling reveals association of PPR proteins with ribosomes in the mitochondria of plants. *Mol Cell Proteomics*.

Said, N., Rieder, R., Hurwitz, R., Deckert, J., Urlaub, H., and Vogel, J. (2009). In vivo expression and purification of aptamer-tagged small RNA regulators. *Nucleic Acids Res*, 37:e133.

Saikawa, N., Akiyama, Y., and Ito, K. (2004). FtsH exists as an exceptionally large complex containing HflKC in the plasma membrane of Escherichia coli. *J Struct Biol*, 146:123–129.

Salvadori, G., Junges, R., Morrison, D. A., and Petersen, F. C. (2019). Competence in Streptococcus pneumoniae and Close Commensal Relatives: Mechanisms and Implications. *Front Cell Infect Microbiol*, 9:94.

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., García-Sotelo, J. S., Alquicira-Hernández, K., Muñiz-Rascado, L. J., Peña-Loredo, P., Ishida-Gutiérrez, C., Velázquez-Ramírez, D. A., Del Moral-Chávez, V., Bonavides-Martínez, C., Méndez-Cruz, C.-F., Galagan, J., and Collado-Vides, J. (2019). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Res*, 47:D212–D220.

Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016). The quantitative and condition-dependent Escherichia coli proteome. *Nat Biotechnol*, 34:104–110.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*, 9:671–675.

Schneider, W. C. (1948). Intracellular distribution of enzymes; the oxidation of octanoic acid by rat liver fractions. *J Biol Chem*, 176:259–266.

Schnorpfeil, A., Kranz, M., Kovács, M., Kirsch, C., Gartmann, J., Brunner, I., Bittmann, S., and Brückner, R. (2013). Target evaluation of the non-coding csRNAs reveals a link of the two-component regulatory system CiaRH to competence control in Streptococcus pneumoniae R6. *Mol Microbiol*, 89:334–349.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, 473:337–342.

Sebert, M. E., Patel, K. P., Plotnick, M., and Weiser, J. N. (2005). Pneumococcal HtrA protease mediates inhibition of competence by the CiaRH two-component signaling system. *J Bacteriol*, 187:3969–3979.

Senkler, J., Senkler, M., Eubel, H., Hildebrandt, T., Lengwenus, C., Schertl, P., Schwarzländer, M., Wagner, S., Wittig, I., and Braun, H.-P. (2017). The mitochondrial complexome of Arabidopsis thaliana. *Plant J*, 89:1079–1092.

Sergeeva, O. V., Prokhorova, I. V., Ordabaev, Y., Tsvetkov, P. O., Sergiev, P. V., Bogdanov, A. A., Makarov, A. A., and Dontsova, O. A. (2012). Properties of small rRNA methyltransferase RsmD: mutational and kinetic study. *RNA*, 18:1178–1185.

Shajani, Z., Sykes, M. T., and Williamson, J. R. (2011). Assembly of bacterial ribo-
    somes. *Annu Rev Biochem*, 80:501–526.

Shanker, E. and Federle, M. J. (2017). Quorum Sensing Regulation of Competence
    and Bacteriocins in Streptococcus pneumoniae and mutans. *Genes*, 8:E15.

Sharpe, P. T. (1988). *Methods of Cell Separation*. Elsevier.

Shchepachev, V., Bresson, S., Spanos, C., Petfalski, E., Fischer, L., Rappsilber, J., and
    Tollervey, D. (2019). Defining the RNA interactome by total RNA-associated pro-
    tein purification. *Mol Syst Biol*, 15:e8689.

Sinha, D., Zimmer, K., Cameron, T. A., Rusch, D. B., Winkler, M. E., and De Lay, N. R.
    (2019). Redefining the Small Regulatory RNA Transcriptome in Streptococcus
    pneumoniae Serotype 2 Strain D39. *J Bacteriol*, 201:e00764–18.

Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007). The RNA chaperone Hfq is
    essential for the virulence of Salmonella typhimurium. *Mol Microbiol*, 63:193–217.

Slager, J., Aprianto, R., and Veening, J.-W. (2018). Deep genome annotation of the
    opportunistic human pathogen Streptococcus pneumoniae D39. *Nucleic Acids Res*,
    46:9971–9989.

Slager, J., Aprianto, R., and Veening, J.-W. (2019). Refining the pneumococcal com-
    petence regulon by RNA-sequencing. *J Bacteriol*, 201:e00780–18.

Smirnov, A., Förstner, K. U., Holmqvist, E., Otto, A., Günster, R., Becher, D.,
    Reinhardt, R., and Vogel, J. (2016). Grad-seq guides the discovery of ProQ as a
    major small RNA-binding protein. *Proc Natl Acad Sci U S A*, 113:11591–11596.

Smirnov, A., Schneider, C., Hör, J., and Vogel, J. (2017a). Discovery of new RNA
    classes and global RNA-binding proteins. *Curr Opin Microbiol*, 39:152–160.

Smirnov, A., Wang, C., Drewry, L. L., and Vogel, J. (2017b). Molecular mechanism
    of mRNA repression in trans by a ProQ-dependent small RNA. *EMBO J*, 36:1029–
    1045.

Son, A., Park, J.-E., and Kim, V. N. (2018). PARN and TOE1 Constitute a 3′ End
    Maturation Module for Nuclear Non-coding RNAs. *Cell reports*, 23:888–898.

Spring, T. G. and Wold, F. (1971). The purification and characterization of Escherichia
    coli enolase. *J Biol Chem*, 246:6797–6802.

Sun, Y. and Vanderpool, C. K. (2013). Physiological consequences of multiple-target
    regulation by the small RNA SgrS in Escherichia coli. *J Bacteriol*, 195:4804–4815.

Sutandy, F. X. R., Hsiao, F. S.-H., and Chen, C.-S. (2016). High throughput platform
    to explore RNA-protein interactomes. *Crit Rev Biotechnol*, 36:11–19.

Sutandy, F. X. R., Qian, J., Chen, C.-S., and Zhu, H. (2013). Overview of protein microarrays. *Curr Protoc Protein Sci*, Chapter 27:Unit 27.1.

Sutherland, C. and Murakami, K. S. (2018). An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase. *EcoSal Plus*, 8:doi:10.1128/ecosalplus.ESP–0004–2018.

Suzuki, S., Tanigawa, O., Akanuma, G., Nanamiya, H., Kawamura, F., Tagami, K., Nomura, N., Kawabata, T., and Sekine, Y. (2014). Enhanced expression of Bacillus subtilis yaaA can restore both the growth and the sporulation defects caused by mutation of rplB, encoding ribosomal protein L2. *Microbiology*, 160:1040–1053.

Svedberg, T. and Pedersen, K. O. (1940). *The Ultracentrifuge.* Oxford: Clarendon Press.

Svedberg, T. and Rinde, H. (1924). The Ultra-Centrifuge, a New Instrument for the Determination of Size and Distribution of Size of Particle in Amicroscopic Colloids. *J Am Chem Soc*, 46(12):2677–2693.

Temin, H. M. and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, 226:1211–1213.

Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., Durkin, A. S., Gwinn, M., Kolonay, J. F., Nelson, W. C., Peterson, J. D., Umayam, L. A., White, O., Salzberg, S. L., Lewis, M. R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A. M., Utterback, T. R., Hansen, C. L., McDonald, L. A., Feldblyum, T. V., Angiuoli, S., Dickinson, T., Hickey, E. K., Holt, I. E., Loftus, B. J., Yang, F., Smith, H. O., Venter, J. C., Dougherty, B. A., Morrison, D. A., Hollingshead, S. K., and Fraser, C. M. (2001). Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science*, 293:498–506.

Thao, S., Chen, C.-S., Zhu, H., and Escalante-Semerena, J. C. (2010). N(epsilon)-lysine acetylation of a bacterial transcription factor inhibits Its DNA-binding activity. *PLoS One*, 5:e15123.

The External RNA Controls Consortium (2005). Proposed methods for testing and selecting the ERCC external RNA controls. *BMC Genomics*, 6:150.

The Tabula Muris Consortium (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, 562:367–372.

The UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*, 47:D506–D515.

The Gene Ontology Consortium (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*, 47:D330–D338.

Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003). Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct*, 32:115–133.

Thomason, M. K., Bischler, T., Eisenbart, S. K., Förstner, K. U., Zhang, A., Herbig, A., Nieselt, K., Sharma, C. M., and Storz, G. (2015). Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. *J Bacteriol*, 197(1):18–28.

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative genomics viewer (igv): high-performance genomics data visualization and exploration. *Briefings Bioinf*, 14:178–192.

Timasheff, S. N. (1993). The control of protein stability and association by weak interactions with water: how do solvents affect these processes? *Annu Rev Biophys Biomol Struct*, 22:67–97.

Tovpeko, Y., Bai, J., and Morrison, D. A. (2016). Competence for Genetic Transformation in Streptococcus pneumoniae: Mutations in SigmA Bypass the ComW Requirement for Late Gene Expression. *J Bacteriol*, 198:2370–2378.

Tree, J. J., Granneman, S., McAteer, S. P., Tollervey, D., and Gally, D. L. (2014). Identification of bacteriophage-encoded anti-sRNAs in pathogenic Escherichia coli. *Mol Cell*, 55:199–213.

Treiber, T., Treiber, N., and Meister, G. (2018). Identification of microRNA Precursor-Associated Proteins. *Methods Mol Biol*, 1823:103–114.

Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiß, J.-L., Eichner, N., Lehmann, G., Schall, K., Urlaub, H., and Meister, G. (2017). A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. *Mol Cell*, 66:270–284.e13.

Trotochaud, A. E. and Wassarman, K. M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. *Nat Struct Mol Biol*, 12:313–319.

Tseng, C.-K., Wang, H.-F., Burns, A. M., Schroeder, M. R., Gaspari, M., and Baumann, P. (2015). Human Telomerase RNA Processing and Quality Control. *Cell Rep*, 13:2232–2243.

Tsui, H.-C. T., Mukherjee, D., Ray, V. A., Sham, L.-T., Feig, A. L., and Winkler, M. E. (2010). Identification and characterization of noncoding small RNAs in Streptococcus pneumoniae serotype 2 strain D39. *J Bacteriol*, 192:264–279.

Ude, S., Lassak, J., Starosta, A. L., Kraxenberger, T., Wilson, D. N., and Jung, K. (2013). Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science*, 339:82–85.

Ueta, M., Ohniwa, R. L., Yoshida, H., Maki, Y., Wada, C., and Wada, A. (2008). Role of HPF (hibernation promoting factor) in translational activity in Escherichia coli. *J Biochem*, 143:425–433.

Ueta, M., Yoshida, H., Wada, C., Baba, T., Mori, H., and Wada, A. (2005). Ribosome binding proteins YhbH and YfiA have opposite functions during 100S formation in the stationary phase of Escherichia coli. *Genes Cells*, 10:1103–1112.

Updegrove, T. B., Zhang, A., and Storz, G. (2016). Hfq: the flexible RNA matchmaker. *Curr Opin Microbiol*, 30:133–138.

Urdaneta, E. C., Vieira-Vieira, C. H., Hick, T., Wessels, H.-H., Figini, D., Moschall, R., Medenbach, J., Ohler, U., Granneman, S., Selbach, M., and Beckmann, B. M. (2019). Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *Nat Commun*, 10:990.

Uzzau, S., Figueroa-Bossi, N., Rubino, S., and Bossi, L. (2001). Epitope tagging of chromosomal genes in Salmonella. *Proc Natl Acad Sci U S A*, 98:15264–15269.

van der Maaten, L. J. and Hinton, G. E. (2008). Visualizing High-Dimensional Data Using t-SNE. *J Mach Learn Res*, 9:2579–2605.

van Dijk, E. L., Jaszczyszyn, Y., and Thermes, C. (2014). Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res*, 322:12–20.

van Opijnen, T. and Camilli, A. (2012). A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res*, 22:2541–2551.

Wadler, C. S. and Vanderpool, C. K. (2007). A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc Natl Acad Sci U S A*, 104:20454–20459.

Wagner, E. G. H. and Romby, P. (2015). Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv Genet*, 90:133–208.

Wang, Y., Cui, T., Zhang, C., Yang, M., Huang, Y., Li, W., Zhang, L., Gao, C., He, Y., Li, Y., Huang, F., Zeng, J., Huang, C., Yang, Q., Tian, Y., Zhao, C., Chen, H., Zhang, H., and He, Z.-G. (2010). Global protein-protein interaction network in the human pathogen Mycobacterium tuberculosis H37Rv. *J Proteome Res*, 9:6665–6677.

Warrier, I., Ram-Mohan, N., Zhu, Z., Hazery, A., Echlin, H., Rosch, J., Meyer, M. M., and van Opijnen, T. (2018). The Transcriptional landscape of Streptococcus pneumoniae TIGR4 reveals a complex operon architecture and abundant riboregulation critical for growth and virulence. *PLoS Pathog*, 14:e1007461.

Wassarman, K. M. (2018). 6S RNA, a Global Regulator of Transcription. *Microbiol Spectr*, 6:RWR–0019–2018.

Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev*, 15:1637–1651.

Wassarman, K. M. and Storz, G. (2000). 6S RNA regulates E. coli RNA polymerase activity. *Cell*, 101:613–623.

Wassarman, K. M., Zhang, A., and Storz, G. (1999). Small RNAs in Escherichia coli. *Trends Microbiol*, 7:37–45.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25:1189–1191.

Watson, J. D. (1965). *Molecular Biology of the Gene.* W. A. Benjamin, Inc. New York.

Weaver, J., Mohammad, F., Buskirk, A. R., and Storz, G. (2019). Identifying Small Proteins by Ribosome Profiling with Stalled Initiation Complexes. *mBio*, 10:e02819–18.

Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D., and Au, K. F. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res*, 6:100.

Wells, J. N., Bergendahl, L. T., and Marsh, J. A. (2016). Operon Gene Order Is Optimized for Ordered Protein Complex Assembly. *Cell Rep*, 14:679–685.

Wöhlbrand, L., Ruppersberg, H. S., Feenders, C., Blasius, B., Braun, H.-P., and Rabus, R. (2016). Analysis of membrane-protein complexes of the marine sulfate reducer Desulfobacula toluolica Tol2 by 1D blue native-PAGE complexome profiling and 2D blue native-/SDS-PAGE. *Proteomics*, 16:973–988.

Wilkinson, M. E., Charenton, C., and Nagai, K. (2019). RNA Splicing by the Spliceosome. *Annu Rev Biochem*, doi:10.1146/annurev-biochem-091719-064225.

Wilton, J., Acebo, P., Herranz, C., Gómez, A., and Amblar, M. (2015). Small regulatory RNAs in Streptococcus pneumoniae: discovery and biological functions. *Front Genet*, 6:126.

Winther, A. R., Kjos, M., Stamsås, G. A., Håvarstein, L. S., and Straume, D. (2019). Prevention of EloR/KhpA heterodimerization by introduction of site-specific amino acid substitutions renders the essential elongasome protein PBP2b redundant in Streptococcus pneumoniae. *Sci Rep*, 9:3681.

Woodford, N. and Livermore, D. M. (2009). Infections caused by Gram-positive bacteria: a review of the global challenge. *J Infect*, 59 Suppl 1:S4–16.

Worhunsky, D. J., Godek, K., Litsch, S., and Schlax, P. J. (2003). Interactions of the non-coding RNA DsrA and RpoS mRNA with the 30 S ribosomal subunit. *J Biol Chem*, 278:15815–15824.

Wuchty, S., Rajagopala, S. V., Blazie, S. M., Parrish, J. R., Khuri, S., Finley, R. L., and Uetz, P. (2017). The Protein Interactome of Streptococcus pneumoniae and Bacterial Meta-interactomes Improve Function Predictions. *mSystems*, 2:e00019–17.

Xue, X., Tomasch, J., Sztajer, H., and Wagner-Döbler, I. (2010). The delta subunit of RNA polymerase, RpoE, is a global modulator of Streptococcus mutans environmental adaptation. *J Bacteriol*, 192:5081–5092.

Yang, Z.-K., Luo, H., Zhang, Y., Wang, B., and Gao, F. (2019). Pan-genomic analysis provides novel insights into the association of E.coli with human host and its minimal genome. *Bioinformatics*, 35:1987–1991.

Zhang, Q., Soares de Oliveira, S., Colangeli, R., and Gennaro, M. L. (1997). Binding of a novel host factor to the pT181 plasmid replication enhancer. *J Bacteriol*, 179:684–688.

Zheng, J. J., Perez, A. J., Tsui, H.-C. T., Massidda, O., and Winkler, M. E. (2017). Absence of the KhpA and KhpB (JAG/EloR) RNA-binding proteins suppresses the requirement for PBP2b by overproduction of FtsA in Streptococcus pneumoniae D39. *Mol Microbiol*, 106:793–814.

Zhou, Z., Sim, J., Griffith, J., and Reed, R. (2002). Purification and electron microscopic visualization of functional human spliceosomes. *Proc Natl Acad Sci U S A*, 99:12203–12207.

Chapter 7

Appendix

7.1   Appendix figures
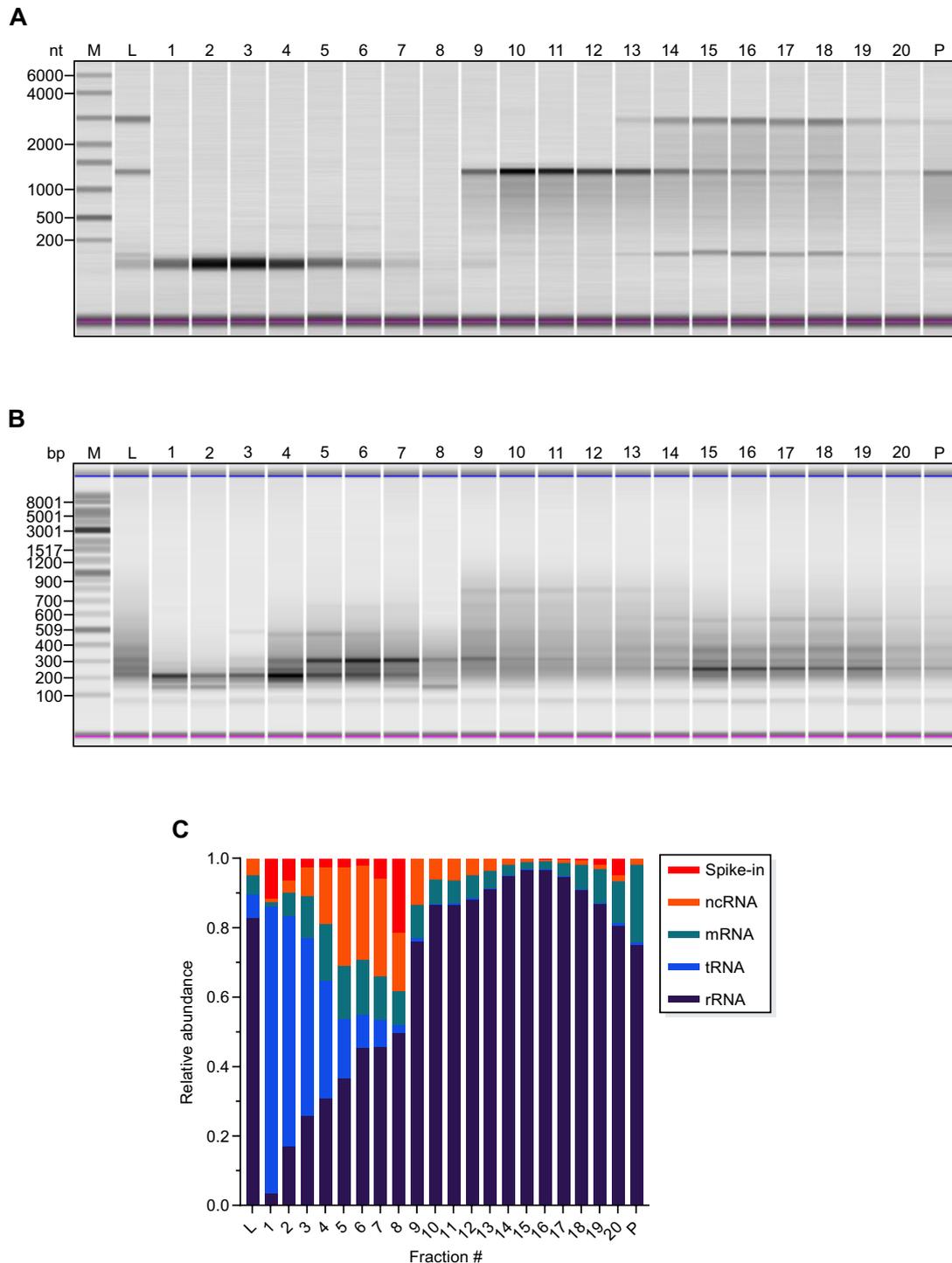
A

B

C

Figure 7.1: Quality control of *E. coli* Grad-seq library preparation and RNA-seq. (**A**) RNA measurements of the purified RNA obtained from an *E. coli* gradient using capillary electrophoresis. The observed migration patterns of the RNA are identical to those shown in Figure 2.4. (**B**) DNA measurements after cDNA library preparation of the RNA shown in (**A**), which was subsequently subjected to sequencing. (**C**) Analysis of the obtained RNA-seq reads reveals that each fraction has a different composition of RNA classes. M, size marker. L, lysate (input control). P, pellet.
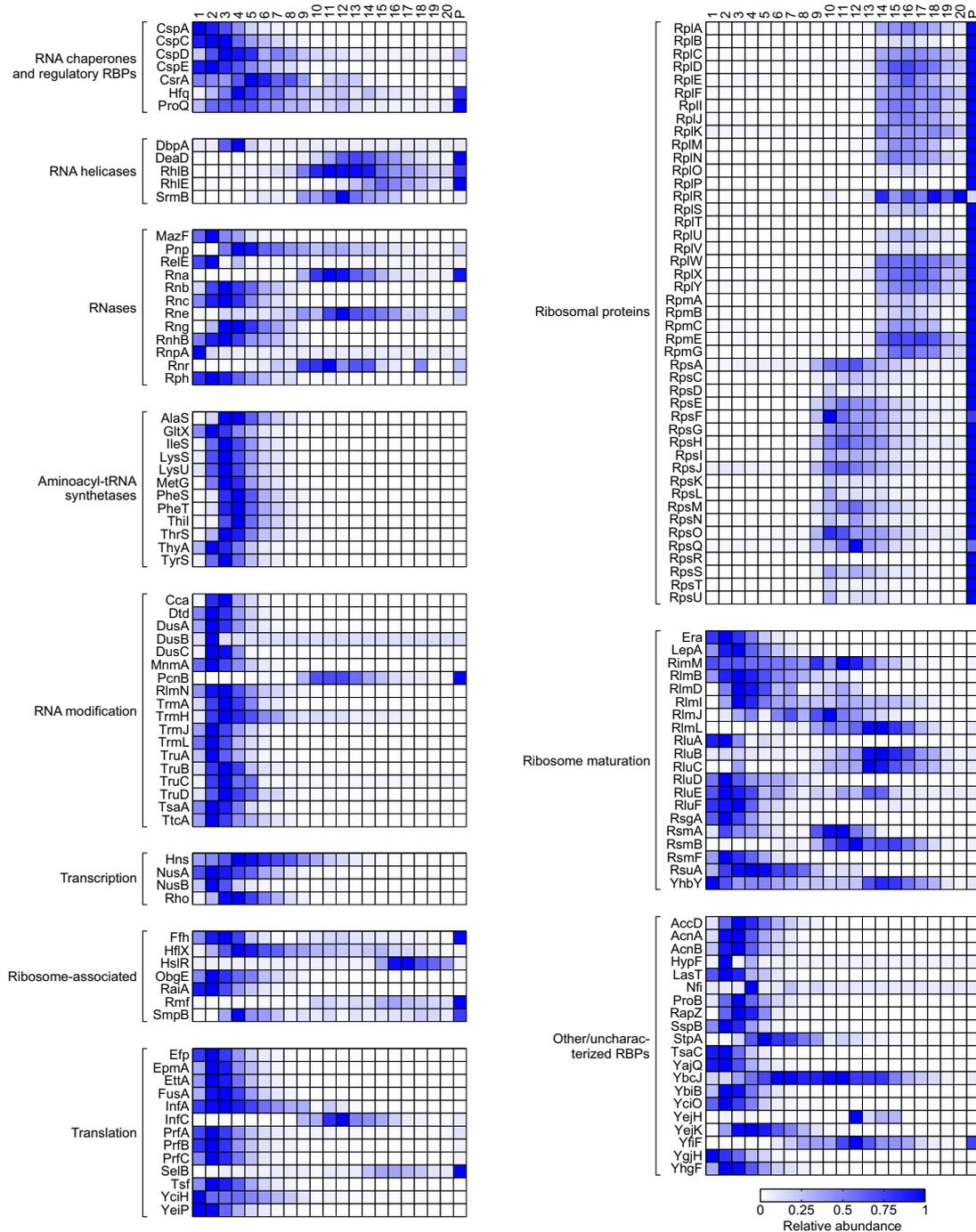
Figure 7.2: Overview of the sedimentation profiles of *E. coli* RBPs. The Grad-seq MS data was filtered for RBPs according to predictions by UniProt (The UniProt Consortium, 2019) and Gene Ontology (Ashburner et al., 2000; The Gene Ontology Consortium, 2019). The RBPs detected in the dataset were subsequently classified according to their functions.
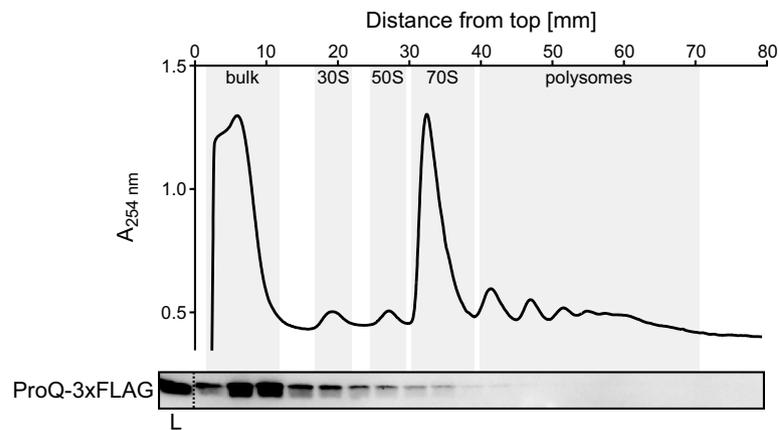
Figure 7.3: Sucrose polysome gradient analysis of ProQ sedimentation. Western blotting of sucrose polysome gradient analysis of a lysate obtained from a *proQ*-3xFLAG strain reveals ribosome association of ProQ. L, lysate (input control).

A

ORF1 **UUGGUUCUGGAGGGGGUUUGUUGUGGGCAAUGA**UGCAUUUAAGUUAUCGUCUGCAGAUAGAGGAGAUAUUACAAUAAACAACGA
AUCAGGGCAUUUGAUAGUCAAUACCGCAAUUCUAUCAGGAGAUAUAGUCACUCUAAGAGGAGGAGAAAUUAGGUUGGUAUUAUA
GCUUGUGCGCGCCAUGAUUGGCGCGCAAUUU

```
LVLEGVCCGQ (10)
```

B

ORF2 UUGGUUCUGG<mark>AGGGG</mark>UUUGUU**GUGGGCAAUGAUGCAUUUAAGUUAUCGUCUGCAGAUAGAGGAGAUAUUACAAUAAACAACGA**
AUCAGGGCAUUUGAUAGUCAAUACCGCAAUUCUAUCAGGAGAUAUAGUCACUCUAAGAGGAGGAGAAAUUAGGUUGGUAUUAUA
GCUUGUGCGCGCCAUGAUUGGCGCGCAAUUU

```
MGNDAFKLSSADRGDITINNESGHLIVNTAILSGDIVTLRGGEIRLVL (48)
```

C

ORF3 UUGGUUC<mark>UGGAGG</mark>GGGUUUG**UUGUGGGCAAUGAUGCAUUUAAGUUAUCGUCUGCAGAUAGAGGAGAUAUUACAAUAA**ACAACGA
AUCAGGGCAUUUGAUAGUCAAUACCGCAAUUCUAUCAGGAGAUAUAGUCACUCUAAGAGGAGGAGAAAUUAGGUUGGUAUUAUA
GCUUGUGCGCGCCAUGAUUGGCGCGCAAUUU

```
MLWAMMHLSYRLQIEEILQ (19)
```

D

ORF4 UUGGUUCUGGAGGGGGUUUGUUGUGGGCAAUGAUGCAUUUAAGUUAUCGUCUGCAGAUAGAGGAGAUAUUACAAUAAACAACGA
AUCAGGGCAUUUGAUAGUCAAUACCGCAAUUCUAUCAGGAGAUAUAGUCACUCUAAGAGGAGGAGAAAUUAGGUUGGUAUUAUA
GCUUGUGCGCGCCAUGAUUGGCGCGCAAUUU

```
MRAFDSQYRNSIRRYSHSKRRRN (23)
```

E

ORF5 UUGGUUCUGGAGGGGGUUUGUUGUGGGCAAUGAUGCAUUUAAGUUAUCGUCUGCAGAUAGAGGAGAUAUUACAAUAAACAACGA
AUCAGGGCAUUUGAUAGUCAAUACCGCAAUUCUAUCAGGAGAUAUAGUCACUCUAAGAGGAGGAGAAAUUAGGUUGGU**AUUAUA**
GCUUGUGCGCGCCAUGAUUGGCGCGCAAUUU

```
MIACARHDWRAI (12)
```

Figure 7.4: Predicted ORFs present in RyeG. (A–E) Prediction of all possible ORFs (highlighted in bold) within RyeG revealed 5 different ORFs. The amino acid sequence of each ORF is given below the corresponding ORF. ORF2 and ORF3 have putative ribosome-binding sites (highlighted in gray) upstream of the start codon.
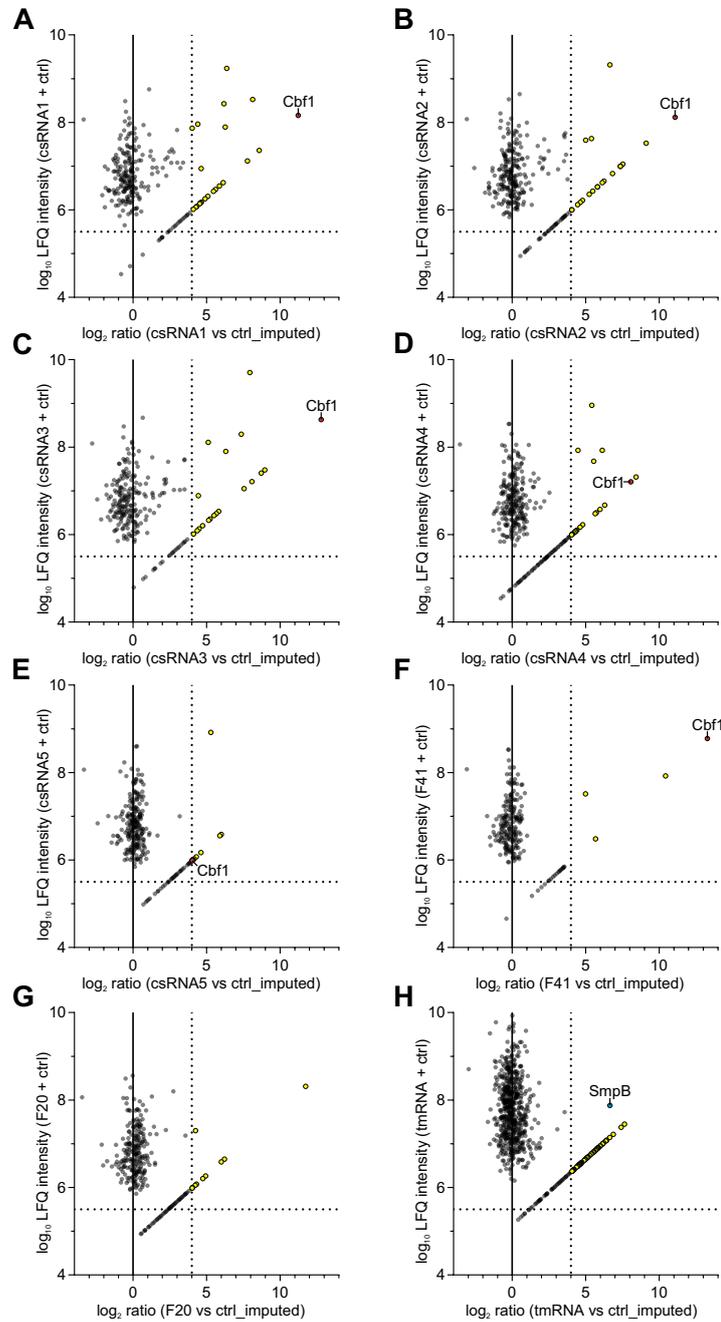
Figure 7.5: MS results of the 14mer pull-downs. (**A–H**) Log_{10} LFQ intensities of the corresponding bait sRNA + the control are plotted against the log_2 ratio between the corresponding sRNA versus the control. Proteins only detected in the pull-down samples were imputed with values close to the baseline in the control (ctrl_imputed) to allow calculation of the ratio. Proteins only detected in the control were not given pseudocounts in the pull-down samples and therefore omitted. Cut-offs for proteins displayed in Figure 3.12 A (highlighted in yellow) were set to 5.5 for the log_{10} LFQ intensities and to 4 for the log_2 ratios (indicated by dotted lines). Cbf1 (highlighted in red) is one of the most abundant and enriched proteins specific for csRNA1–5 (**A–E**) and F41 (**F**), whereas it was not pulled down by the riboswitch RNA F20 (**G**) or tmRNA (**H**). The positive control tmRNA enriched its specific binding protein SmpB (highlighted in blue). Proteins not considered specifically enriched are shown in gray.

Figure 7.6: Multiple sequence alignment of Cbf1. Multiple sequence alignment of Cbf1 homologs from different members of the Firmicutes. The tRNA_anti and HD domains are highlighted based on their predicted positions within *S. pneumoniae* Cbf1 (El-Gebali et al., 2019; Mitchell et al., 2019). Residues with ≥50 % identity are highlighted in a blue gradient. Alignment was executed using Clustal Omega (Madeira et al., 2019) and visualized using Jalview (Waterhouse et al., 2009).
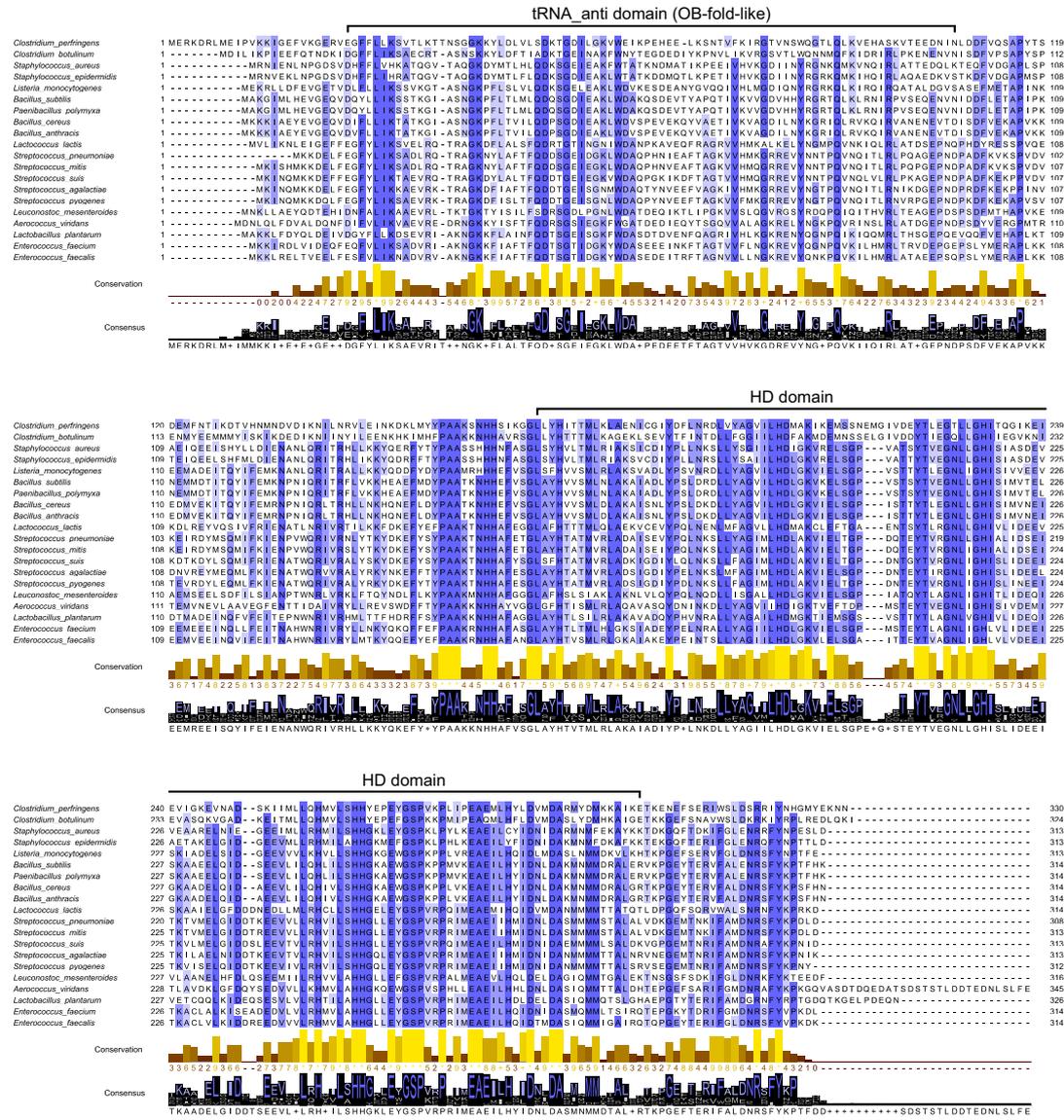
Figure 7.7: Multiple sequence alignment of SP_2226. Multiple sequence alignment of SP_2226 homologs from different members of the Firmicutes. Residues with ≥50 % identity are highlighted in a blue gradient. Alignment was executed using Clustal Omega (Madeira et al., 2019) and visualize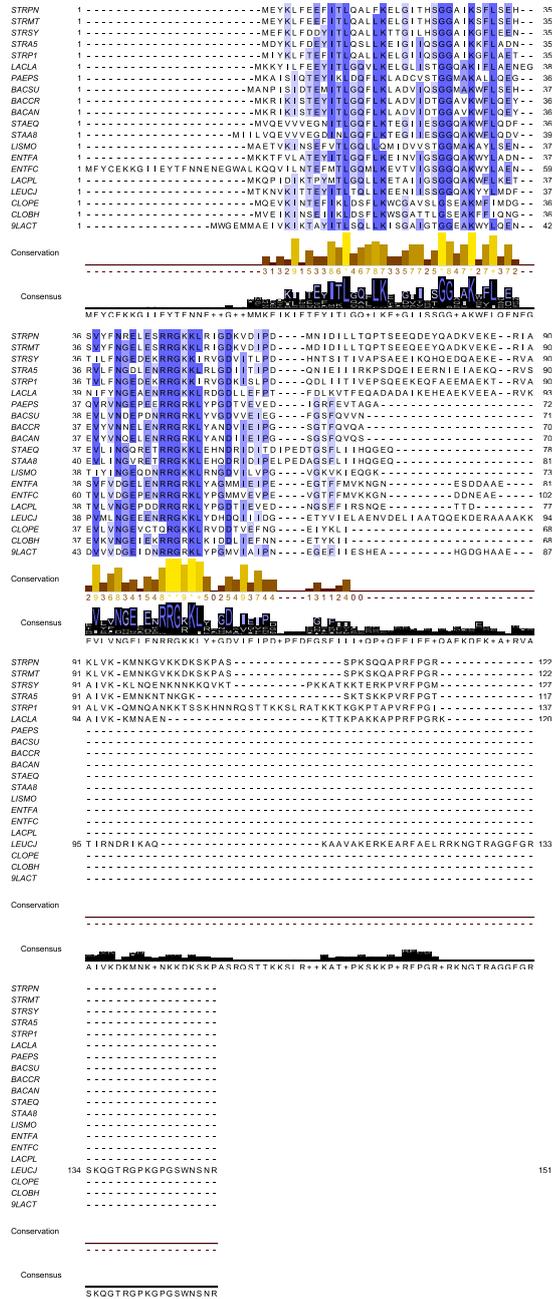d using Jalview (Waterhouse et al., 2009). STRPN, *Streptococcus pneumoniae*. STRMT, *Streptococcus mitis*. STRSY, *Streptococcus suis*. STRA5, *Streptococcus agalactiae*. STRP1, *Streptococcus pyogenes*. LACLA, *Lactococcus lactis*. PAEPS, *Paenibacillus polymyxa*. BACSU, *Bacillus subtilis*. BACCR, *Bacillus cereus*. BACAN, *Bacillus anthracis*. STAEQ, *Staphylococcus epidermidis*. STAA8, *Staphylococcus aureus*. LISMO, *Listeria monocytogenes*. ENTFA, *Enterococcus faecalis*. ENTFC, *Enterococcus faecium*. LACPL, *Lactobacillus plantarum*. LEUCJ, *Leuconostoc carnosum*. CLOPE, *Clostridium perfringens*. CLOBH, *Clostridium botulinum*. 9LACT, *Aerococcus viridans*.

7.2   List of abbreviations

Table 7.1: List of abbreviations.

Abbreviation	Meaning
2D-MS	2D gel analysis followed by MS
aa	amino acid
AD	activating domain
AP/MS	affinity purification followed by MS
APS	ammonium persulfate
BD	binding domain
bp	base pair
cDNA	complementary DNA
CDS	coding sequence
CFU	colony forming unit
CLASH	UV-crosslinking, ligation and sequencing of hybrids
CLIP-seq	crosslinking immunoprecipitation followed by RNA-seq
csBN-MS	cryo-slicing blue native-MS
CSP	competence stimulating peptide
csRNA	cia-dependent sRNA
ctrl	control
DNA	deoxyribonucleic acid
DNAP	DNA polymerase
DNase	deoxyribonuclease
dNTP	deoxyribonucleotide
dRNA-seq	differential RNA-seq
dsRNA	double-stranded RNA
DTT	dithiothreitol
EAL	ethanolamine ammonia-lyase
EDTA	ethylene diamine tetraacetic acid
ERCC	External RNA Control Consortium
fCLIP-seq	formaldehyde CLIP-seq
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
gDNA	genomic DNA
GradR	RNase-sensitive gradient profiles
Grad-seq	gradient profiling by sequencing
HD	histidine/aspartate
L	lysate
LB	Lennox broth
LC-MS/MS	liquid chromatography-tandem MS
M	marker
MBP	maltose binding protein
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger RNA
MS	mass spectrometry
NB	northern blot
ncRNA	noncoding RNA
nt	nucleotide
OB	oligonucleotide/oligosaccharide-binding
$OD_{600\,nm}$	optical density at $600\,nm$
ORF	open reading frame
P	pellet

Table 7.1, *continued.*

Abbreviation	Meaning
P/C/I	phenol/chloroform/isoamyl alcohol
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis
PCP	protein correlation profiling
PCR	polymerase chain reaction
PNPase	polynucleotide phosphorylase
POI	protein of interest
PPI	protein-protein interaction
RBP	RNA-binding protein
RBS	ribosome binding site
R-DeeP	RNA-dependent proteins
RIL-seq	RNA interaction by ligation and sequencing
RIP-seq	RNA immunoprecipitation followed by RNA-seq
RNA	ribonucleic acid
RNAP	RNA polymerase
RNase	ribonuclease
RNA-seq	RNA sequencing
RNP	ribonucleoprotein particle
rRNA	ribosomal RNA
RT-qPCR	reverse transcription quantitative PCR
SD	Shine-Dalgarno
SD	standard deviation
SDS	sodium dodecyl sulfate
sRNA	small regulatory RNA
SRP	signal recognition particle
TA	toxin/antitoxin
TCS	two-component system
TEMED	N,N,N',N'-tetramethylethylenediamine
TF	transcription factor
tmRNA	transfer-messenger RNA
tRNA	transfer RNA
t-SNE	t-stochastic neighbor embedding
TSS	transcriptional start site
UAS	upstream activating sequence
UPS	universal protein spike-in
UTR	untranslated region
v/v	volume/volume
w/v	weight/volume
WB	western blot
wt	wild type
Y2H	yeast two-hybrid

7.3 Curriculum vitae

7.4 Publications

7.4.1 Published manuscripts

Hör, J., Garriss, G., Di Giorgio, S., Hack, L.-M., Vanselow, J. T., Förstner, K. U., Schlosser, A., Henriques-Normark, B., and Vogel, J. (2020). Grad-seq in a Gram-positive bacterium reveals exonucleolytic sRNA activation in competence control. *EMBO J*, doi:10.15252/embj.2019103852

Hör, J., Matera, G., Vogel, J., Gottesman, S., and Storz, G. (2020). Trans-Acting Small RNAs and Their Effects on Gene Expression in *Escherichia coli* and *Salmonella enterica*. *EcoSal Plus*, doi:10.1128/ecosalplus.ESP-0030-2019

Hör, J., Gorski, S. A., and Vogel, J. (2018). Bacterial RNA Biology on a Genome Scale. *Mol Cell*, 70:785–799

Smirnov, A., Schneider, C., **Hör, J.**, and Vogel, J. (2017). Discovery of new RNA classes and global RNA-binding proteins. *Curr Opin Microbiol*, 39:152–160

Hör, J. and Vogel, J. (2017). Global snapshots of bacterial RNA networks. *EMBO J*, 36:245–247

Dawson, W., **Hör, J.**, Egert, M., van Kleunen, M., and Pester, M. (2017). A Small Number of Low-abundance Bacteria Dominate Plant Species-specific Responses during Rhizosphere Colonization. *Front Microbiol*, 8:975

7.4.2 Manuscripts in preparation

Hör, J., Gerovac, M., Di Giorgio, S., Förstner, K. U., and Vogel, J.; A Grad-seq map of RNA and protein complexes discovers new stable interactions with ribosomes in *Escherichia coli*.

Hör, J. and Vogel, J.; Analysis of the RNA and protein complexome by Grad-seq.

Acknowledgments

Affidavit/Eidesstattliche Erklärung

I hereby confirm that my thesis entitled "Discovery of RNA/protein complexes by Grad-seq" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date   Signature

Hiermit erkläre ich an Eides statt, die Dissertation "Ermittlung von RNA/Protein-Komplexen mittels Grad-seq" eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ort, Datum   Unterschrift