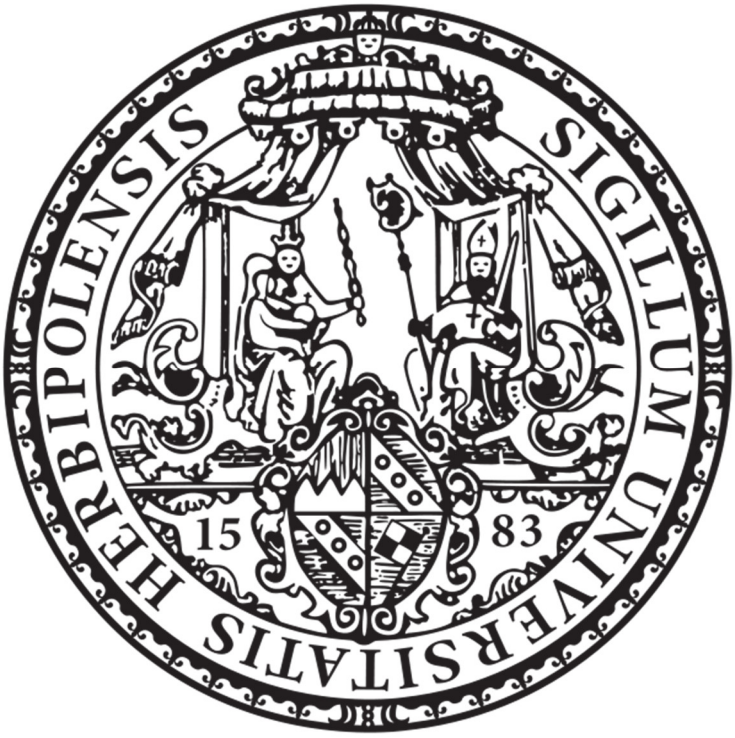
Small RNA-binding complexes in *Chlamydia trachomatis* identified by Next-Generation Sequencing techniques



Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von

Maximilian Andreas Klepsch
Koblenz

Würzburg, 2019

Eingereicht am: …….....

Mitglieder der Prüfungskommission:

Vorsitzender: ……..

Gutachter: Univ.-Prof. Dr. Thomas Rudel

Gutachter: Univ.-Prof. Dr. Thomas Dandekar

Tag des Promotionskolloquiums: …….....

Doktorurkunde ausgehändigt am: …….....

Eidesstattliche Erklärungen nach §7 Abs. 2 Satz 3, 4, 5 der Promotionsordnung der Fakultät für Biologie

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation: „**Identifizierung von kleinen RNA-bindenden Komplexen in *Chlamydia trachomatis* mittels Hochdurchsatz-Sequenziertechniken**“, 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.

Weiterhin erkläre ich, dass bei allen Abbildungen und Texten bei denen die Verwertungsrechte (Copyright) nicht bei mir liegen, diese von den Rechtsinhabern eingeholt wurden und die Textstellen bzw. Abbildungen entsprechend den rechtlichen Vorgaben gekennzeichnet sind sowie bei Abbbildungen, die dem Internet entnommen wurden, der entsprechende Hypertextlink angegeben wurde.

Affidavit

I hereby declare that my thesis entitled: „**Small RNA-binding complexes in *Chlamydia trachomatis* identified by Next-Generation Sequencing techniques**” 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 verify that the thesis has not been submitted as part of another examination process neither in identical nor in similar form.

Besides I declare that if I do not hold the copyright for figures and paragraphs, I obtained it from the rights holder and that paragraphs and figures have been marked according to law or for figures taken from the internet the hyperlink has been added accordingly.

Würzburg, den_____

_____
Signature PhD-student

Abstract

Chlamydia infect millions worldwide and cause infertility and blinding trachoma. *Chlamydia trachomatis* (*C. trachomatis*) is an obligate intracellular gram-negative pathogen with a significantly reduced genome. This bacterium shares a unique biphasic lifecycle in which it alternates between the infectious, metabolically inert elementary bodies (EB) and the non-infections, metabolically active replicative reticular bodies (RB).

One of the challenges of working with *Chlamydia* is its difficult genetic accessibility. In the present work, the high-throughput method TagRNA-seq was used to differentially label transcriptional start sites (TSS) and processing sites (PSS) to gain new insights into the transcriptional landscape of *C. trachomatis* in a coverage that has never been achieved before. Altogether, 679 TSSs and 1067 PSSs were detected indicating its high transcriptional activity and the need for transcriptional regulation. Furthermore, the analysis of the data revealed potentially new non-coding ribonucleic acids (ncRNA) and a map of transcriptional processing events. Using the upstream sequences, the previously identified σ^{66} binding motif was detected.

In addition, Grad-seq for *C. trachomatis* was established to obtain a global interactome of the RNAs and proteins of this intracellular organism. The Grad-Seq data suggest that many of the newly annotated RNAs from the TagRNA-seq approach are present in complexes. Although *Chlamydia* lack the known RNA-binding proteins (RBPs), e.g. Hfq and ProQ, observations in this work reveal the presence of a previously unknown RBP.

Interestingly, in the gradient analysis it was found that the σ^{66} factor forms a complex with the RNA polymerase (RNAP). On the other hand, the σ^{28} factor is unbound. This is in line with results from previous studies showing that most of the genes are under control of σ^{66}. The ncRNA IhtA is known to function via direct base pairing to its target RNA of HctB, and by doing so is influencing the chromatin condensation in *Chlamydia*. This study confirmed that lhtA is in no complex. On the other hand, the ncRNA ctrR0332 was found to interact with the SNF2 protein ctl0077, a putative helicase. Both molecules co-sedimented in the gradient and were intact after an aptamer-based RNA pull-down. The SWI2/SNF2 class of proteins are nucleosome remodeling complexes. The prokaryotic RapA from *E. coli* functions as transcription regulator by stimulating the RNAP recycling. This view might imply that the small ncRNA (sRNA) ctrR0332 is part of the global regulation network in *C. trachomatis* controlling the transition between EBs and RBs via interaction with the SNF2 protein ctl0077.

The present work is the first study describing a global interactome of RNAs and proteins in *C. trachomatis* providing the basis for future interaction studies in the field of this pathogen.

Zusammenfassung

Chlamydien verursachen jährlich Millionen Neuinfektionen weltweit und können zu Spätschäden wie Unfruchtbarkeit und Erblindung führen. Chlamydien sind obligat intrazelluläre, gram-negative Pathogene mit einem stark reduzierten Genom. Sie besitzen einen einzigartigen biphasischen Lebenszyklus, bei dem der Erreger zwischen den metabolisch inaktiven, infektiösen Elementarkörperchen (EBs) und den nicht infektiösen, metabolisch aktiven und replikativen Retikularkörperchen (RBs) alterniert.

Eine Problemantik beim Arbeiten mit Chlamydien ist die Schwierigkeit der gezielten genetischen Manipulation des Pathogens.

In der vorliegenden Arbeit wurde die Hochdurchsatz-Sequenziermethode TagRNA-Seq genutzt, um die transkriptionelle Organisation von *Chlamydia trachomatis* (*C. trachomatis*) zu analysieren und besser zu verstehen. Transkriptionelle Start Stellen (TSS) und Prozessierungsstellen (PSS) werden dabei unterschiedlich markiert, sodass eine zuverlässigere und genauere Auflösung erreicht wird als bisher durch in anderen Studien verwendete Methoden. Insgesamt konnten so 679 TSSs und 1067 PSSs detektiert werden. Es konnte gezeigt werden, dass das Transkriptom von *C. trachomatis* weitaus aktiver ist als bisher angenommen und eine Regulation auf transkriptioneller Ebene bedarf. Die Methode erlaubte zudem die Identifizierung von potenziell neuen nicht-kodierende RNAs sowie die Kartierung von transkriptionellen Prozessierungsereignissen. Unter Verwendung der 5'-upstreamliegenden Sequenzen konnte außerdem das in anderen Bakterien bereits bekannte σ^{66}-Bindemotiv detektiert werden.

In der vorliegenden Arbeit wurde zudem die Methode Grad-Seq in *C. trachomatis* etabliert, um ein globales Interaktom für RNAs (*engl. ribonucleic acid)* und Proteine des intrazellulären Organismus zu erstellen.

Für viele der im TagRNA-Seq Ansatz identifizierten und neu annotierten RNAs konnte so eine Komplexbildung beobachtet werden. Dies deutet auf das Vorhandensein eines bislang unbekanntes RNA-Bindeprotein (RBP) hin, da Chlamydien keines der bekannten RBPs, z.B. Hfq oder ProQ, besitzen.

Die Gradienten-Analyse ergab, dass der σ^{66}-Faktor in einem Komplex mit der RNA-Polymerase (RNAP) vorliegt und dass der σ^{28}-Faktor ungebunden ist. Diese Beobachtung entspricht den Ergebnissen vorheriger Studien, die zeigten das die meisten Gene durch σ^{66} kontrolliert werden. Die Daten bestätigen außerdem, dass IhtA, eine ncRNA (*engl*. non-coding ribonucleic

acid), die über direkte Basenpaarbindung mit ihrem Ziel-RNA von *hctB* interagiert, nicht in einem Komplex vorliegt. Für die ncRNA ctrR0332 hingegen konnte das SNF2-Protein ctl0077 als Interaktionspartner identifiziert werden. Beide Moleküle co-sedimentieren im Gradienten und konnten mittels eines Aptamer-basierenden RNA Pull-Downs in intakter Form isoliert werden. Die Klasse der SWI2/SNF2-Proteine gehört zu den Nukleosomen-Remodeling-Komplexen. In Prokaryoten konnte für das in *E. coli* vorkommende RapA, welches ebenfalls zu den SWI2/SNF2-Proteinen zählt, die Funktion eines Transkriptionsregulators nachgewiesen werden, indem die RNAP-Wiederverwertung stimuliert wird. Dies könnte bedeuten, dass die ncRNA ctrR0332 ebenfalls Teil eines globalen Regulationsnetzwerks ist, welches durch Interaktion mit dem SNF2-Protein ctl0077 die Transition zwischen dem RB- und EB-Stadium reguliert.

In der vorliegenden Arbeit konnte erstmals ein globales Interaktom von RNAs und Proteinen in *C. trachomatis* erstellt werden, welches als Grundlage für zukünftige Interaktionsstudien des Pathogens genutzt werden kann.

Table of Contents

1. Introduction

1.1. Chlamydia

Chlamydia are gram negative obligate intracellular pathogens within a wide range of host species. In 1907, Halberstaedter and his colleague Prowazek showed that the symptoms of trachoma are experimentally transferable from humans to apes. Using Giemsa staining, they identified intracytoplasmic vacuoles in the cells of conjunctival smear and described them as the cause of trachoma. These organisms were described as intermediates between virus and bacteria and the name Chlamydozoa was proposed (Halberstädter and Prowazek, 1907). The word *Chlamydiae* derived from the Greek word “chlamys”, which is an ancient Greek term for cloak. In the years after the original discovery, similar inclusion bodies were found in new-borns suffering from conjunctivitis, in the cervical smear of their respective mothers as well as in male urethral swab (Halberstädter and Prowazek, 1907). In the 1960s, *Chlamydiae* were classified as bacteria with the help of molecular methods and the development of electron microscope and ultramicrotome (Moulder, 1966).

Originally, the family of *Chlamydiaceae* was composed of two species, *C. trachomatis* and *Chlamydia psittaci* (*C. psittaci*). This was based upon their glycogen acquisition in the inclusion (Gordon and Quan, 1965) and their resistance to sulfadiazine (Lin and Moulder, 1966). In the years to follow, new organisms were found, e.g. *Chlamydia pneumoniae* (*C. pneumoniae)* (Grayston et al., 1986), *Chlamydia pecorum* (Fukushi and Hirai, 1992) and chlamydia-like bacteria e.g. *Simkania negevensis* (Kahane et al., 1995), and *Parachlamydia acanthamoebae* (Amann et al., 1997).

With advancing DNA (Deoxyribonucleic acid) techniques and the need for revision of the phylogenetic *Chlamydia* tree, Everett and colleagues used 16S and 23S rRNA analysis to redefine the taxonomy of the *Chlamydia*. *Chlamydia* is now the order followed by the family *Chlamydiaceae* which is in return composed of the two genera of *Chlamydophila* and *Chlamydia* (Everett et al., 1999)(Figure 1).

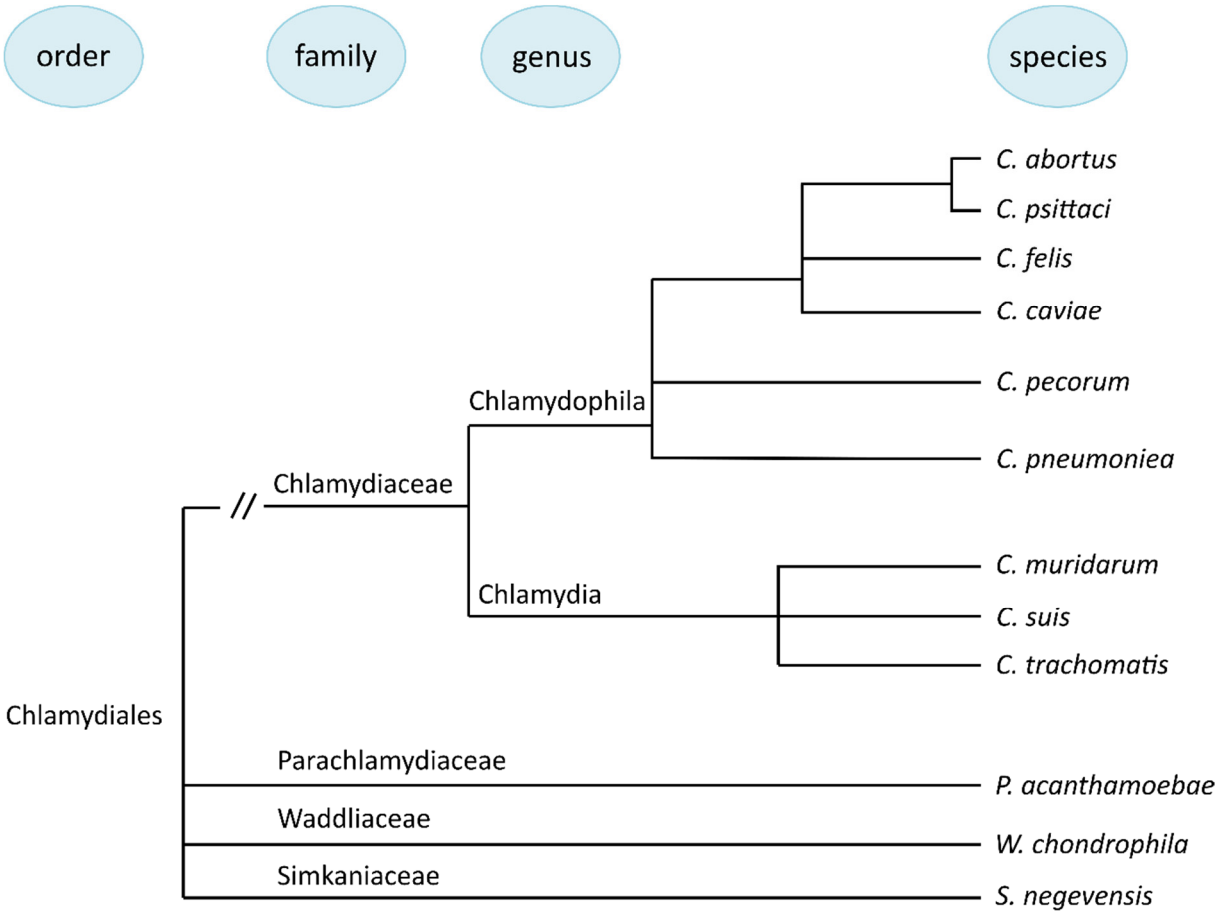


Figure 1: Taxonomy of the order *Chlamydiales*. The tree is based on ribosomal DNA sequence data. The order of *Chlamydiales* is composed of four families. The *Chlamydiaceae* family includes two genera containing nine species in total. Figure modified after (Bush and Everett, 2001).

1.2. Development and lifecycle

Chlamydia have a unique biphasic lifecycle, which was first described by Bedson and Bland, who analysed the replication of a causative agent (Bedson and Bland, 1932). Electron microscopy showed two morphological distinct forms of *Chlamydia*; the elementary bodies (EBs) and the reticulate bodies (RBs) (Constable, 1959, Gaylord, 1954). EBs are spore-like particles with a diameter of 0.3 µm (Eb et al., 1976). This form is extracellular, able to infect susceptible cells and, until recently, has been assumed to be metabolically inactive, since the chromatin at this stage is highly condensed by the histone-like proteins HctA and HctB (Brickman et al., 1993). Recent work indicated that EBs show metabolic and biosynthetic activity (Omsland et al., 2012). Initially, EBs attach to the cell by electrostatic interactions (Kuo et al., 1972, Kuo et al., 1973). In a second step of adhesion several *Chlamydia* membrane proteins are suggested to play a role including outer membrane proteins OmpA, outer membrane complex B protein OmcB or the polymorphic outer membrane protein (pmp) (Grimwood et al., 2001). The current hypothesis is that the OmpA of *Chlamydia* and heparan

sulphate on the cell side interact with each other (Zhang and Stephens, 1992, Swanson and Kuo, 1991, Su et al., 1996). While the electrostatic interaction is reversible, the second step of adhesion is irreversible and to date not completely understood (Carabeo and Hackstadt, 2001).

There are several hypotheses circulating about the EB internalization after adhesion, however, the exact mode of internalisation of the EBs is still not understood in its entirety (Scidmore-Carlson and Hackstadt, 2000). Receptor mediated endocytosis is the predominating hypothesis (Wyrick et al., 1989, Hodinka et al., 1988). A variety of host receptors were found, e.g. heparin sulphate, mannose receptor, mannose 6-phosphate receptor and the estrogen receptor (Puolakkainen et al., 2005, Kuo et al., 2002, Davis et al., 2002). Alternative hypotheses postulate that *Chlamydia* enter the cell via clathrin-independent endocytosis or that lipid rafts are involved in their entry (Jutras et al., 2003, Stuart et al., 2003). After adhesion, the *Chlamydia* effector protein TARP (translocated actin recruiting phosphoprotein) is secreted into the host cell (Subtil et al., 2000). TARP gets phosphorylated and initiates actin polymerisation with Rho GTPases (Jewett et al., 2006, Carabeo et al., 2004, Clifton et al., 2004). In the following, several host tyrosine kinases are activated (Kim et al., 2011, Fawaz et al., 1997, Elwell et al., 2008, Birkelund et al., 1994) leading to cytoskeletal rearrangements and eventually to the uptake of the EBs into the non-phagocytic cells (Dautry-Varsat et al., 2005, Carabeo, 2011). The internalized EBs are located in membrane-bound vesicles derived from the endolysosomal pathway. These vesicles are called inclusions and *Chlamydia* will stay in these compartments for the rest of its developmental cycle (Moulder, 1991). The vesicles bear no markers for the lysosomal pathway and block fusion with endosomes and lysosomes (Hackstadt, 2000, Al-Younes et al., 1999). Within the first eight hours post infection (hpi), the EBs start to differentiate to RBs. The RBs are metabolically active, bigger in size (1 µm) and undergo binary fission for their replication. At this stage, the *Chlamydia* has an inner and outer membrane, like gram-negative bacteria, and are non-infectious. After several divisions, the binary fission becomes asynchronous (McClarty, 1994, Moulder, 1991). Most of the RBs begin to differentiate into EBs, followed by their release either by cell lysis or extrusion. These released EBs infect the surrounding cells. The regulation of the RB-EB transition is not completely understood, but it has been hypothesized that factors like space limitation play a role (Hackstadt et al., 1997, Bavoil et al., 2000). On a molecular level, the DNA condensation is regulated by histone-like proteins HctA (Hc1) and HctB (Hc2), which are expressed during

the late phase of the RBs and are thought to initiate condensation of the DNA to the state observed in EBs (Perara et al., 1992, Hackstadt et al., 1991).

Several stimuli can cause *Chlamydia* to go into persistence. This state can be triggered by stress factors, such as interferon gamma (IFN-γ) (Shemer and Sarov, 1985), viral co-infection (Vanover et al., 2008, Deka et al., 2006), tryptophan starvation (Beatty et al., 1994) and antibiotics like penicillin (Skilton et al., 2009). During these stimuli, *Chlamydia* differentiates into aberrant bodies (ABs). These ABs are, similar to the RBs, not infectious but can be transferred via cell division. ABs are larger in size due to multiple copies of the genome and can re-differentiate back to RBs after removal of the stimuli (Byrne and Ojcius, 2004, Moulder et al., 1980)(Figure 2).

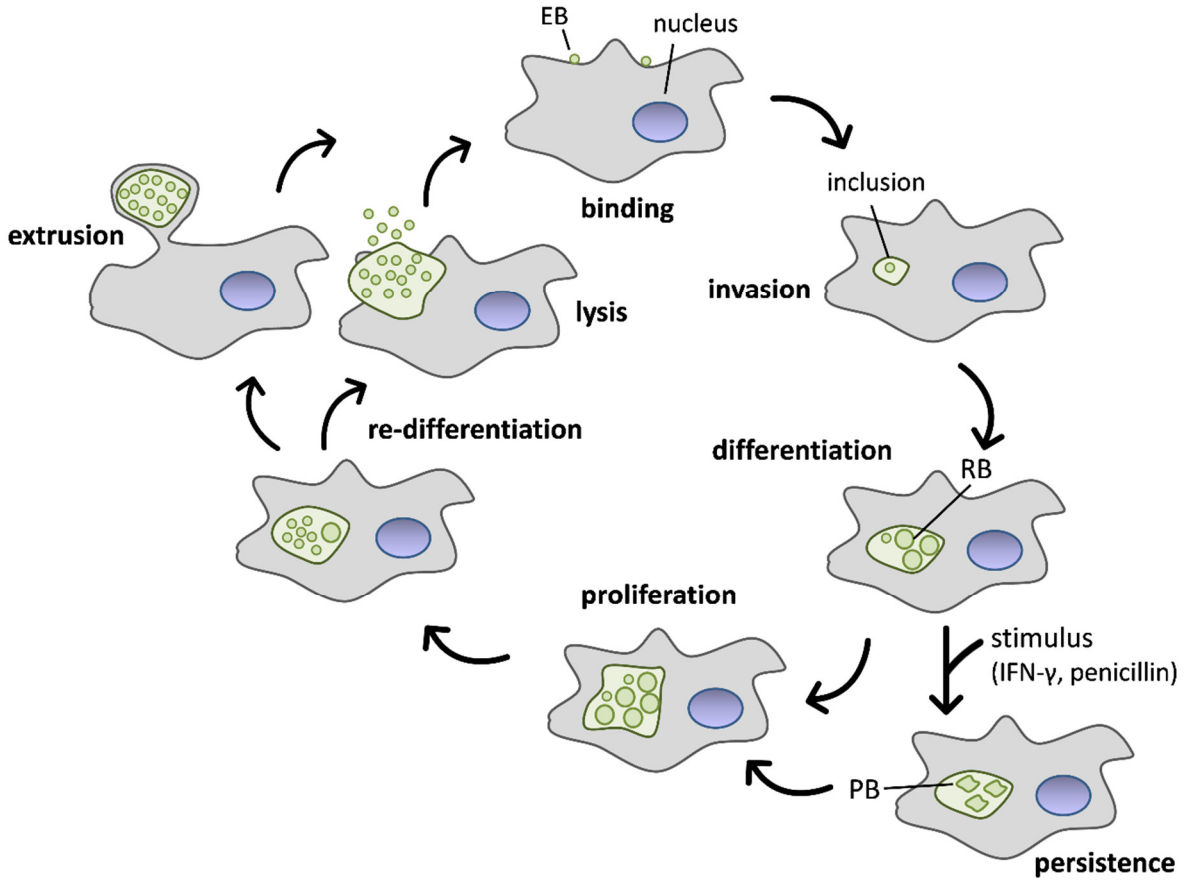


Figure 2: *C. trachomatis* lifecycle. Infectious elementary bodies (EBs) attach to the host cell. The EB is endocytosed into a membrane-bound compartment, the inclusion. EBs than differentiate into reticulate bodies (RBs) followed by intracellular growth and division of RBs in the inclusion. After re-differentiation of RBs into EBs, EBs are released from the host cell by lysis or extrusion. Under stress conditions, RBs enter a persistent state and differentiate into enlarged persistent bodies (PBs). Reactivation of the bacteria can be achieved by removal of the stress stimuli. Figure modified after (Byrne and Ojcius, 2004).

1.3. *Chlamydia trachomatis* pathogenesis

Under the whole *Chlamydiaceae* family, *C. trachomatis* and *C. pneumoniae* are the two major species infecting humans.

C. pneumoniae is a pathogen causing respiratory infections like acute pneumonia, bronchitis, sinusitis and pharyngitis (Kern et al., 2009). It accounts for up to 10% of community acquired pneumonia (Elwell et al., 2016). Besides respiratory diseases, *C. pneumoniae* was also linked to atherosclerotic processes (Kuo et al., 1993).

C. trachomatis is the most common cause of sexually transmitted diseases worldwide with around 131 million estimated new infections annually (WHO, 2019). *C. trachomatis* strains are divided into three biovars including the trachoma biovar, the genital tract biovar and the lympho granuloma venerum (LGV) biovar. The biovars are further subdivided by serovar. The trachoma biovar, which includes the serovars A-C, is the causative agent of a recurrent disease affecting the conjunctiva and cornea of the eye and the leading cause of non-congenital blindness primarily in developing nations (Elwell et al., 2016).

Sexually transmitted infections of the urogenital tract are caused by *C. trachomatis* serovars D-K. In females, the infection manifests as endocervicitis (Malhotra et al., 2013). 70-80% of female infections are asymptomatic or only have slight symptoms and therefore are not treated and can ascend to the upper genital tract (Elwell et al., 2016). Non-treatment of these infections can thus lead to salpingitis, endometritis, pelvic inflammatory disease, ectopic pregnancy and tubal factor infertility (Malhotra et al., 2013). New-borns can get infected during birth via an affected birth passage. In most cases, these new-borns than suffer from conjunctivitis, and more rarely from otitis media (RKI, 2010). Genital infections in males first manifest as urethritis with symptoms like pain when urinating(RKI, 2010). Since up to 50% of men also have asymptomatic infections, which are not treated, the infections can ascend to the prostate and further into the epididymis (Malhotra et al., 2013, RKI, 2010). Chlamydial infections causing prostatitis are also discussed to result in male infertility (Stojanov et al., 2018, RKI, 2010).

The serovars L1-L3 cause LGV, a sexually transmitted infection mainly occurring in the tropics, manifesting as an inflammation in the lymph nodes (RKI, 2010). Infection with *C. trachomatis* is also associated with an increased risk of HIV (human immunodeficiency virus) transmission and with the development of cervical carcinoma (Malhotra et al., 2013).

The Centres for Disease Control and Prevention recommends treating chlamydial infections with azithromycin or doxycycline. Alternate regimens include, amongst others, erythromycin or ofloxacin. Altogether, the treatment is dependent on different factors including the site of infection, age of the patient, an existing pregnancy and complexity of infection (Malhotra et al., 2013, CDC, 2015). Until now, there is no protective vaccination available to control chlamydial infections, despite different attempts to develop one (Malhotra et al., 2013).

1.4. Chlamydial Genomes

Chlamydial genomes are about 1 Mb in size with about 850-1100 genes (Stephens et al., 1998). Due to the relatively small genome size, *Chlamydia* were prime candidates for whole-genome sequencing. In 1998, the *C. trachomatis* serovar D genome was published (Stephens et al., 1998). In the years that followed, several other *Chlamydia* were sequenced including *C. pneumoniae* CWL029 (Kalman et al., 1999), *C. muridarum* Nigg (Read et al., 2000), *C. pneumoniae* AR039 (Read et al., 2000) and *C. pneumoniae* J138 (Shirai et al., 2000). Up to date, advances in next-generation sequencing enabled the assembly of over 170 genomes for *C. trachomatis*, 68 genomes for *C. psittaci* and 28 genomes for *C. muridarum*. The obligatory intracellular lifestyle of *Chlamydia* reflects itself in their genomes which all are highly reduced in size and show a reduced gene content in comparison to other bacteria e.g. *E.coli* (Trevors, 1996). The evolution of the chlamydial genomes or any prokaryotic genomes can be explained by six fundamental processes (as summarised by (Koonin and Wolf, 2008)): 1) genome streamlining, 2) genome degradation, 3) complexification by gene duplication, 4) operon shuffling, 5) complexification via horizontal gene transfer and 6) propagation of mobile elements (Koonin and Wolf, 2008). In most pathogens genomic simplification and reduction are associated with genomic degradation (Moran, 2002). In *Chlamydia,* it appears that the genomic reduction and simplification are the result of genomic streaming (Sigalova et al., 2019). *Chlamydia* only harbour up to 15 predicted transcription factors (Domman and Horn, 2015). Due to the high number of genome assemblies for this genus, pan-genomic analyses have been performed, revealing that chlamydial species only possess a low number of pseudogenes (Bachmann et al., 2014, Nunes and Gomes, 2014). Furthermore, it was shown that the genome is highly conserved outside the plasticity zone, which was shown to be located in the regions of ~160 kb (*C. pneumoniae*) and ~50 kb (*C. trachomatis*) around the termination origin, where a higher rate of DNA-reorganisation occurs (Read et al., 2000). The

Gene content across *Chlamydia* species is conserved, and the majority is shared across the phylum (Collingro et al., 2011). *Chlamydia* lack destructive mobile elements with the exceptions of fragments of IS (insertion)-like elements, e.g. the IS-associated tetracycline resistance in *Chlamydia suis*, and prophages in a variety of chlamydial genomes (Nunes and Gomes, 2014, Sachse et al., 2014, Vorimore et al., 2013, Sachse et al., 2015). Altogether, these genome are characterised by streamlining and are generally highly adapted, yet they provide a basis for bacteria of the genus *Chlamydia* to explore broad host specificities and tissue tropisms (Sachse et al., 2014, Vorimore et al., 2013, Sachse et al., 2015).

1.5. Sigma-factors and transcriptional regulation

The chlamydial RNA transcriptional apparatus consists, similar to other eubacteria, of the α, β and β' core RNA apparatus and the exchangeable σ(sigma)-factors (Koehler et al., 1990, Gu et al., 1995, Engel et al., 1990). σ-factors play a crucial role in transcription initiation in bacteria und are used to adjust the transcriptome in response to different stimuli e.g. stress or different environmental signals. The major σ-factor is σ^{66}, encoded by the gene RNA polymerase sigma factor *rpoD*. In addition, *Chlamydia* are using two other factors for their gene expression, σ^{28}, encoded by the gene *rpoD*, and σ^{54} encoded by the gene *rpoD* (Kalman et al., 1999, Stephens et al., 1998). Alternative σ-factors play a crucial role in the regulation of virulence genes in other bacteria (Kazmierczak et al., 2005, Fang, 2005). In *Chlamydia,* it was proposed that the temporal gene expression is under control of alternative σ-factors, even before they were found (Plaunt and Hatch, 1988, Engel et al., 1990, Fahr et al., 1995). The two alternative σ-factors have homologs in other bacteria, e.g. *Bacillus subtilis* (Stephens et al., 1998). In *Chlamydia,* the σ-factors are not adjacent to their regulatory genes, as in *Bacillus subtilis*, instead they are spread across the genome (Stephens et al., 1998, Kalman et al., 1999). Both factors have orthologs in all *Chlamydia,* except σ^{28} is missing in *Protochlamydiaceae* (Horn et al., 2004).

In other bacteria, it is shown that the σ^{54}-factor is regulating nitrogen metabolism (Merrick, 1993) and the σ^{28}-factor regulates the genes of flagellar synthesis, chemotaxis and motility (Haldenwang, 1995). The alternative σ-factor in *Chlamydia* recognizes a subset of genes in the late lifecycle of *Chlamydia* (Yu and Tan, 2003, Yu et al., 2006). Interestingly, late expressed genes are always either expressed by the σ^{66} RNA polymerase (RNAP) or the σ^{24} RNAP (Yu and Tan, 2003, Fahr et al., 1995, Mathews et al., 1993). Yu and colleges showed that the histon-like

proteins HctB and HctA are either σ^{24} or σ^{66} regulated, which might suggest that the RB-EB transition is controlled by more than just one regulatory networks (Yu and Tan, 2003), since these proteins condense the DNA during RB-EB-transition (Brickman et al., 1993).

The σ^{24}-factor itself appears to be regulated by alternative means. The σ-protein is already expressed during the middle of the chlamydial lifecycle, but the genes expressed by the σ-factor, such as HctB, are present in the late stage of the chlamydial lifecycle (Shen et al., 2004, Douglas and Hatch, 2000).

In *Bacillus subtilis* (*B. subtilis*), the anti-σ-factor RsbW binds to the σ-factor σ^{24} so that the σ^{B} RNAP is inhibited (Price, 2002). This has been called partner-switching mechanism, because the RsbW binds either to the σ-factor or to its anti-anti-σ-factor RsbV (Price, 2002). When RsbV binds to RsbW, the σ-factor is released, and the RNAP is functional (Price, 2002). RsbV can bind to RsbW only in its unphosphorylated state, while RsbW can phosphorylate RsbV. RsbU, an upstream regulator, can dephosphorylate RsbV (Price, 2002).

Chlamydia possess RsbV orthologs, two proteins named RsbV1 and RsbV2, which are sequence-similar to RsbV of *B. subtilis* and two orthologs of RsbU from *B. subtilis*, CT589 and RsbU (Stephens et al., 1998, Hua et al., 2006).

The current proposed model for σ^{24}-factor in *Chlamydia* is that the σ-factor is expressed at 16 hpi (Shen et al., 2004), but its anti-σ-factor RsbW is already expressed at 3 hpi (Belland et al., 2003), hindering premature expression of late genes under σ^{24}-factor influence. The two anti-anti-σ-factors are expressed during the mid-phase genes (Belland et al., 2003), which would allow σ^{24} to regulate its late gene expression. RsbU and CT589 could use their predicted extracytoplasmic region to sense external stimuli and regulate the σ-factor via parallel pathways.

Late genes, which are not under control of σ^{24}, like *omcAB* and *hctA*, are expressed by the σ^{66} RNAP, but this RNAP also expresses genes in the early and midcycle (Fahr et al., 1995). There is evidence that EUO functions as a repressor for late genes, which are expressed by the σ^{66} RNAP, as it is expressed at 1 hpi with *Chlamydia* and is bound by the promotor region of the *omcAB* operon (Wichlan and Hatch, 1993). EUO binds to the AT-rich sequences (Zhang et al., 2000, Zhang et al., 1998) and selectively represses late promotors (Rosario and Tan, 2012). It is still unclear, how EUO is released from its binding sites. During the infection cycle, the EUO protein levels decrease and the protein is absent during the late phase or EBs (Zhang et al., 1998). Zhang and colleagues showed that in *C. psittaci* EUO transcripts were only present at 1

hpi to 20 hpi with a maximum at 15 hpi but could not explain how the EUO levels are controlled during the chlamydial lifecycle (Zhang et al., 1998).

In contrast to the EUO, the levels of Scc4 increase during the late lifecycle of *Chlamydia* (Rao et al., 2009a). It was proposed that Scc4 inhibits the σ^{66} RNAP and facilitates the σ^{24} RNAP (Rao et al., 2009a). Scc4 also acts as a potential σ^{66} RNAP inhibitor in EBs, where its level is more abundant than in RBs (Rao et al., 2009a). Moreover, Scc4 was proposed to function as a chaperon for the type III secretion system and to interact with another type III chaperon (SccI) (Spaeth et al., 2009, Betts-Hampikian and Fields, 2010).

It was long proposed that EBs are transcriptionally inactive, although the σ^{66} RNAP is present (Skipp et al., 2005, Sixt et al., 2011, Shaw et al., 2002). The respective messenger RNA (mRNA) was found in EBs but is most likely produced in RBs at the late stage before transition of RBs into EBs (Belland et al., 2003, Maurer et al., 2007). Furthermore, the translation in extracellular EBs was recently described (Haider et al., 2010). The two histone-like proteins HctA and HctB, which are only transcribed in the late phase by either the σ^{66} RNAP (Fahr et al., 1995) or later by the σ^{28} RNAP (Yu and Tan, 2003), are proposed to inhibit the overall transcription and translation by chromatin condensation (Pedersen et al., 1996, Pedersen et al., 1994, Barry et al., 1993). While HctA binds to supercoiled DNA and creates compacted nucleoid (Barry et al., 1992, Brickman et al., 1993), HctB binds to RNA and linear DNA (Pedersen et al., 1995) and is not present in environmental *Chlamydiae* (Collingro et al., 2011). Finally, it has been proposed that transcription is regulated by the supercoiling-responsive promoters of histone-like proteins (Barry et al., 1993).

1.6. Classes of non-coding RNAs and gene regulation by non-coding RNAs in bacteria

Besides alternative σ-factors and transcription factors, non-coding RNAs (ncRNAs) play a crucial role in bacterial gene regulation with profound effects on cell physiology (Gottesman, 2005). There are different types of ncRNAs including ribosomal RNAs(rRNAs), transfer RNAs (tRNA) and small RNAs (sRNAs). sRNAs are the largest and best-studied group of ncRNAs. In general, sRNAs are around 100 nucleotides in length, specifically controlling the expression of targets via different strategies (Gottesman, 2005). They act as regulators on multiple levels of gene expression including translation, transcription, DNA maintenance or silencing, and mRNA stability (Ahmed et al., 2016). sRNAs either bind to proteins and affect their activities or they

act by base-pairing with their target mRNA leading to formation of RNA duplexes (Repoila and Darfeuille, 2009). Interactions of sRNAs with mRNA can be divided into two groups: cis-encoded sRNAs and trans-encoded sRNAs (Figure 3).

Cis-encoded RNAs or true antisense sRNAs are a class of sRNAs that share extensive complementarity with their target mRNAs. Generally, the sRNA and its target are encoded by the same region of the DNA but are transcribed from opposite strands of the gene they regulate (Tan and Bavoil, 2012). Most of the cis-encoded sRNAs are involved in the control of replication, conjugation and stability of plasmids, in the regulation of transposition of transposons as well as in the fine-tuning of the decision between lysis or lysogeny in bacteriophages (Brantl, 2012b). Besides residing on plasmids or other mobile genetic elements, some cis-encoded sRNAs were found to have chromosomal origin (Waters and Storz, 2009). These sRNAs act as antitoxins and are involved in the alteration of metabolism under stress conditions (Brantl, 2002). There are different regulatory mechanisms employed by cis-encoded sRNAs with direct blocking of the ribosome binding site being one of the simplest and most commonly used. This process results in translation inhibition and has been found to be involved in the control of plasmid replication and maintenance and to be used, amongst others, in the FinOP repressor systems (Brantl, 2007, Koraimann et al., 1996). Here, expression of *traJ*, which is an activator of plasmid F, is controlled by the antisense RNA FinP and RNA-binding protein (RBP) FinO (Gubbins et al., 2003). FinO protects FinP from decay and promotes duplex formation between FinP and *traJ* mRNA (Jerome et al., 1999). Binding of FinP to *traJ* sequesters the *traJ* ribosome binding site leading to prevention of translation and repression of plasmid transfer (Koraimann et al., 1996).

In 2007, Brantl extensively reviewed other known regulatory mechanisms, such as transcription attenuation, inhibition of primer attenuation, inhibition of pseudoknot formation, mRNA degradation, and mRNA stabilization (Brantl, 2007).

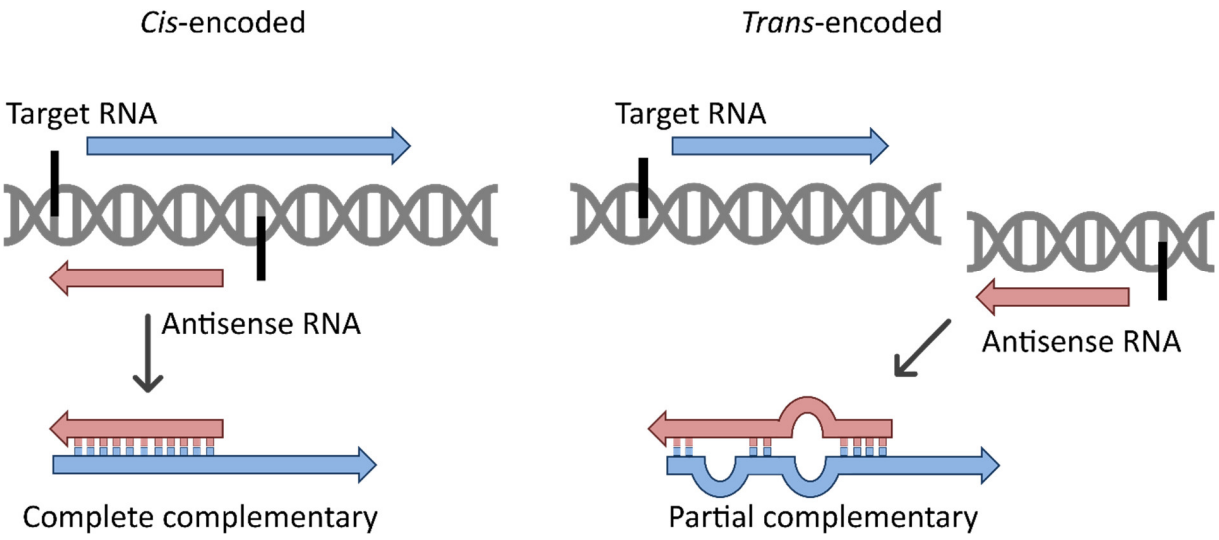


Figure 3: Overview of cis- and trans-encoded small RNAs. Cis-encoded sRNAs are encoded by the same region of the DNA as their target mRNA and share extensive complementarity with it, whereas trans-encoded sRNAs are encoded at a different genomic location than its target and share only limited complementarity with it. Antisense RNAs are displayed in red and sense RNAs in blue. Black rectangles indicate promoters. Figure modified after (Brantl, 2012a).

In contrast to the group of cis-encoded sRNAs, trans-encoded sRNAs are encoded at different genomic locations than their targets and share limited complementarity with them. Trans-encoded RNAs are capable of multiple base pairing interactions since the pairing regions are only around 10-25 nt long. Thus, partial and discontinuous patches between sRNA and target mRNA are made (Waters and Storz, 2009). Trans-encoded sRNAs can act negatively as well as positively. Many sRNAs inhibit translation by binding and blocking the ribosome binding site (Gottesman and Storz, 2011). Other sRNAs bind their target mRNAs and stimulate translation initiation by preventing the formation of an inhibitory secondary structure. These alterations of the ribosome binding site structure facilitate access by the translational machinery (Gottesman and Storz, 2011, Waters and Storz, 2009, Tan and Bavoil, 2012) Most of the trans-encoded sRNAs are expressed in response to specific environmental conditions like iron limitation (RhyB), oxidative stress (OxyS), low temperature (DsrA) and others (Repoila and Gottesman, 2003, Masse and Gottesman, 2002, Masse et al., 2003, Altuvia et al., 1998, Altuvia et al., 1997).

Pairing of sRNAs with their target mRNA often requires the RNA chaperone Hfq because of limited complementarity between sRNA and target mRNA (Gottesman and Storz, 2011, Updegrove et al., 2015).

As mentioned above, sRNAs can also interact with proteins and change their activities. sRNAs are mimicking other nucleic acids and compete with their natural counterparts for binding positions to the protein (Gottesman and Storz, 2011, Tan and Bavoil, 2012). The two

best-characterized examples of regulatory sRNAs that act in this way are the 6S RNA from *Escherichia coli* (*E. coli)*, mimicking a DNA promotor open complex and interacting with the RNAP, and the CsrA (carbon storage regulator)/RsmA (ribosomal small subunit methyltransferase A) family regulators that act by titrating their target away from their mRNA-binding sites (Wassarman, 2007, Duss et al., 2014).

In the last years, another class of RNA regulator was found in bacteria, the so-called CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) RNAs. Together with a group of associated proteins, the CRISPR RNAs provide bacteria acquired resistance to bacteriophages and plasmids by recognition and degradation of exogenous DNA (Perez-Reytor et al., 2016, Ahmed et al., 2016). The CRISPR arrays are comprised of conserved short DNA repeat sequences originating from foreign DNA that are separated by unique spacers (Waters and Storz, 2009). The CRISPR arrays are transcribed as long RNAs, which are processed to crRNAs (Sorek et al., 2008). Through base pairing of these sRNAs with foreign nucleic acid, degradation is initiated (Sorek et al., 2008).

1.7. RNA-binding proteins in bacteria

As mentioned above, sRNAs are modulators of gene expression, but their action needs the presence of helper proteins. In bacteria, three RNA-binding proteins (RBP) – CsrA, ProQ and Hfq – are known.

CsrA, the carbon storage regulator, is a translational regulator that is present in many bacteria affecting different biological processes such as biofilm formation, carbon metabolism and quorum sensing (Sabnis et al., 1995, Lenz et al., 2005, Jackson et al., 2002). The CsrA protein binds to the ribosome binding site region of mRNA thus preventing gene expression (Majdalani et al., 2005). The CsrB RNA competes with the target mRNA for CsrA protein binding (Majdalani et al., 2005). If CsrB RNA is present, CrsA is preferentially bound by CsrB, so that the ribosome binding site is not sequestered by CsrA and translation is started (Majdalani et al., 2005).

ProQ is a chromosomal homologue of FinO and was originally identified as an osmoregulatory factor that is needed for regulation of the proline channel protein ProP (Kunte et al., 1999, Chaulk et al., 2011). However, ProQ is now known to target more than 70 sRNAs (Smirnov et al., 2017). Its function is not only to stabilize these sRNAs, but also to promote RNA duplex formation (Chaulk et al., 2011, Smirnov et al., 2017). ProQ preferentially binds transcripts that

are characterized by a complex structure indicating that ProQ targets a distinct class of RNAs compared to CsrA and Hfq which both bind to single-stranded RNAs (Smirnov et al., 2016b).

As aforementioned, the RNA chaperone Hfq is a major player in gene regulation mediated by sRNAs that act via limited base pairing. Hfq can be found in many bacterial species in which it promotes fast base-pairing of sRNA and target mRNA allowing an immediate response when intra- or extracellular conditions change (Wagner, 2013). Hfq was identified around 50 years ago as a host-factor for bacteriophage Qβ replication in *E. coli*. (Franze de Fernandez et al., 1968). Proteins within the Hfq family are thermostable and compromise between 70 and 110 amino acids (Brennan and Link, 2007). Hfq forms a homohexameric doughnut-shaped structure with two binding sites, one proximal and one distal, with both having distinct binding preferences. The distal site of Hfq preferentially binds poly(A) tracts, whereas the proximal site was found to bind AU-rich sequences in single stranded regions (Schumacher et al., 2002, Olejniczak, 2011, Link et al., 2009). In addition, a third binding site has been identified, the lateral site, which is responsible for binding the sRNA body (Sauer et al., 2012). Since there are different binding sites, simultaneous binding of sRNA and mRNA to Hfq is possible promoting the interaction between these two (Van Assche et al., 2015).

The Hfq homohexamer is similar to the eukaryotic Sm and Sm-like proteins, that are central components of the splicing machinery and therefore also involved in post-transcriptional regulation (Vogel and Luisi, 2011). There are different aspects of how Hfq can facilitate base pairing including facilitation of sRNA to the 5`-end of the target mRNA, thus inhibiting translation (Vogel and Luisi, 2011). In return, Hfq can guide sRNA to disruptive secondary structures in the 5'-end of the mRNA enabling translation (Vogel and Luisi, 2011). Another mode of action is the protection of sRNAs from ribonucleases as RNA cleavage is prevented when they are bound by Hfq (Vogel and Luisi, 2011). Hfq can also facilitate the 3`-5`-exonucleolytic degradation by presenting the 3`-end of the RNA for polyadenylation (Vogel and Luisi, 2011).

In *Chlamydia,* Hfq or other RNA-binding proteins are not known. Nevertheless, the previous identification of sRNAs in *Chlamydia* strongly suggests that RNA-binding proteins play a role in chlamydial gene regulation.

1.8. Small RNAs in *Chlamydia*

Not much is known about sRNAs in *Chlamydia* as only a few previous studies focused on them. Interestingly, none of the canonical RBPs, like Hfq, ProQ or CsrA, are encoded on the chlamydial genomes. Therefore, it is interesting how the small RNAs present in *Chlamydia* would function.

One of the known sRNAs in *Chlamydia* is IhtA (inhibitor of *hctA*). IhtA is a trans-encoded sRNA, which was found to inhibit the translation of *hctA* (Grieshaber et al., 2006a, Grieshaber et al., 2006b). Interestingly, it was shown that this sRNA functions by direct base paring to its target RNA (Tattersall et al., 2012). HctB is not affected by this sRNA, indicating the specificity of IhtA (Grieshaber et al., 2006a, Grieshaber et al., 2006b). It was also shown that IhtA is expressed during the early stage of the chlamydial lifecycle and that the protein levels of HctA are low in RBs in the presence of IhtA (Grieshaber et al., 2006a, Grieshaber et al., 2006b). This finding also would make IhtA a regulatory factor in addition to the previously discussed temporal regulatory systems.

With the advent of Next-generation sequencing technologies, two complete chlamydial genomes (*C. trachomatis* and *C. pneumoniae*) were screened for sRNAs (Albrecht et al., 2011, Albrecht et al., 2010). Several new sRNAs were found in *C. trachomatis* and *C. pneumoniae* with a high variation of abundance depending on the life stage of the two *Chlamydia* species (Albrecht et al., 2011, Albrecht et al., 2010). One of the sRNAs re-annotated as ctrR0332 (previously annotated as CTLon_0332) is located downstream of *ltuB* (Albrecht et al., 2011, Albrecht et al., 2010). Interestingly, this RNA makes up 78% of the previously published EB transcriptome and 20% of the total transcripts in RBs. Hence, it appeared to be the most abundant transcript in *C. trachomatis* and *C. pneumoniae* (Albrecht, 2011) and homologs for this sRNA could only be found in *Chlamydia* spp. (Albrecht et al., 2011, Albrecht et al., 2010). Due to varying abundance throughout the chlamydial lifecycle, due to its chlamydial specificity and due to the lack of known RNA chaperons, this small RNA could be another major regulator in the chlamydial system. These previous results indicate, that the chlamydial transcriptome is regulated on several levels with the exact modes of action remaining unknown.

1.9. Current state of Next-generation sequencing

There were two fundamental findings laying the foundation for the development of Next-generation Sequencing: the discovery of the DNA´s structure and the introduction of

methods allowing to determine the nucleic acid sequence (Wu and Kaiser, 1968, Watson and Crick, 2003, Sanger et al., 1977, Maxam and Gilbert, 1977). The first direct sequencing approach was done by Wu and Kaiser using DNA polymerase to determine the 12 bp nucleotide sequence of the cohesive ends of bacteriophage lamda (Wu and Kaiser, 1968). In the years to follow, Maxam and Gilbert used a chemical procedure for DNA sequencing (Maxam and Gilbert, 1977). With the development of dideoxynucleotide chain termination in the mid-1970s, Frederick Sanger and his colleagues revolutionized the sequencing area (Sanger et al., 1977). With earlier methods being time consuming, work intensive and error prone, the Sanger sequencing technique was faster and more efficient in these years (Th and Ma, 2015).

Although Sanger sequencing, which is considered to be the “first generation” of DNA sequencing, is still a useful application for targeted sequencing today, the method has restricted applications because of technical limitations (Rizzo and Buck, 2012). The main bottle neck is sequencing volume or throughput, meaning the amount of DNA fragments that can be sequenced at a time (Rizzo and Buck, 2012). Second-generation sequencing technologies overcomes these limitations. Second-generation sequencing technology is massively parallel, sequencing millions of fragments simultaneously in a short period of time. Also known under the name “high throughput” sequencing, the biggest advantages of second-generation sequencing technologies over Sanger sequencing are the speed, the lower costs and the sequencing output (Kchouk et al., 2017).

Although many sequencing platforms are available, Illumina is currently the market leading company in the sequencing industry. Illumina´s sequencing workflow includes four steps. In library preparation, the DNA is fragmented and specific adapters on both ends are ligated allowing hybridization to the flow cell surface. Cluster generation occurs on a flow cell, a glass slide with lanes that are randomly coated with oligos that are complementary to library adapters. Through bridge amplification, each fragment is amplified into clonal clusters. After cluster generation, the next step is to determine the nucleotide sequence. For this, Illumina uses the Sequencing by synthesis (SBS)-technology. Fluorescently labelled dNTPs and other sequencing reagents are added, and the first base is incorporated. The method employs reversible terminators to detect single bases directly after incorporation during each cycle. Excited by a laser, each of the four DNA bases emits an intensity of a unique wavelength for

identification. The cycle is repeated until acquired read length is achieved. After data acquisition the generated short reads are analysed using different bioinformatic tools.

There are two sequencing read types, single end sequencing, which involves sequencing the DNA from only one end, whereas paired-end sequencing allows to sequence both ends of the fragment (Illumina, 2019, Illumina, 2013, Illumina, 2015).

Besides Illumina, there are two other major sequencing platforms, Roche 454 and Ion torrent sequencing. Roche 454 uses pyrosequencing technique that is based on pyrophosphate detection to report whether a particular base was incorporated in the growing DNA strand, whereas in the Ion torrent semiconductor sequencing technology a hydrogen ion is released when a correct nucleotide is incorporated leading to a change in pH (Kchouk et al., 2017, Slatko et al., 2018).

In the last decade, NGS technologies were continuously improved so that today, the application of second-generation sequencing has a broad range, from the analysis of a few genes using gene panels to exome sequencing or whole genome sequencing. Furthermore, additional applications are transcriptome analysis using RNA-sequencing (Voelkerding et al., 2009) or DNA methylation analysis (Barros-Silva et al., 2018).

Second-generation sequencing technologies for transcriptome analysis are commonly used to compare gene expression profiles of organisms. Furthermore, it can be used to annotate the transcript boundaries by enriching primary transcripts through digestion of 5’polyphosphatetranscripts. This technique is called dRNA-seq and gives global maps of transcriptional start sites (TSSs) (Sharma et al., 2010). Another approach, which is known under the name TaqRNA-seq, uses differential labelling of transcripts by their 5’RNA ends to identify not only TSSs, but also processing sites (PSSs) (Innocenti et al., 2015). By using a glycerol gradient to separate complexes from an organism by their density and analysing the occurrence of RNA molecules by high-throughput sequencing technologies throughout the gradient, insights into RNA complexes and RNA interaction partners can be obtained (Rederstorff et al., 2010). In Grad-Seq, this approach is extended by mass spectrometry to measure the distribution of proteins in the gradient so that new RNA-protein complexes can be found (Smirnov et al., 2016).

However, all second-generation technologies only produce short reads making it difficult to analyse complex genomic regions as repetitive sequences or structural variations (Pollard et al., 2018). Another source of misrepresentation of reads in second-generation data comes

through PCR bias during sample preparation, which leads to a distortion of the sequencing result (Kebschull and Zador, 2015).

Several of these limits can be overcome by third-generation sequencing, which is also known under the name long-read sequencing. The idea of third-generation sequencing is to produce long fragments exceeding several kilobases using easy sample preparation without PCR amplification at lower costs and in less time than second-generation sequencing (Kchouk et al., 2017).

The currently available approaches of third-generation sequencing are the Single molecule real time sequencing approach (SMRT) that is commercialized by Pacific Biosciences and Nanopore Sequencing that was released by Oxford Nanopore Technologies in the form of the MinION Instrument. SMRT technology uses a single DNA polymerase immobilized in a well called zero-mode waveguides. Here, the incorporation of fluoresce labelled nucleotides can be measured in Realtime (Shendure et al., 2017). On average, reads with a length of 10-30 kb can be produced but can exceed over 80 kb (van Dijk et al., 2018). Nanopore sequencing measures the ionic current when the DNA is passing through a single pore embedded in a membrane (Kasianowicz et al., 1996). Nanopore sequencing has the potential to generate ultra-long reads over 800kb (Jain et al., 2018)

Both technologies generate long reads overcoming the problems of short reads and are beneficial for repetitive sequences or structural variant identification, making these technologies prime candidates for genome assembly. However, both techniques are prone to higher error rates compared to second-generation sequencing techniques (Korlach, 2013, Laver et al., 2015).

1.10. Aim of the work

C. trachomatis is a prevalent human pathogen causing urethritis and blindness. Since the genetical manipulation of *C. trachomatis* is still in early stages and is unreliably, the aim of the present work was to resort to global screens to analyse the interactome of *C. trachomatis* and to further deepen the knowledge of the pathogen's transcriptome. Therefore, Grad-Seq was used to separate the complexes of *C. trachomatis* by density in a glycerol gradient. Using high-throughput methods for each fraction of the glycerol gradient will help to determine which molecules are potentially interacting with one another or which molecules form complexes. Additionally, TagRNA-seq was performed to gain further knowledge about the transcriptomic landscape of *C. trachomatis*. Here, RNA species are differentially labelled by their 5' end to differentiate between TSS and PSS. Combining these methods will gain new insights about small RNAs in *Chlamydia* and the potential RBPs as well as insights in the regulation of the Chlamydial transcriptome.

2. Material and Methods

2.1. Material

2.1.1. Bacterial Strains

Experiments in the present work were performed using *C. trachomatis* L2/434/bu, which was obtained from ATCC® VR-902B™. Different *E. coli* strains were used for molecular work (Table 1).

Table 1: *E. coli* strains used in this work.

E. coli strain	Usage
E. coli DH5α	cloning
E. coli Xl1blue	cloning

2.1.2. Cell lines

Propagation and cultivation of *C. trachomatis* was performed in Hela229 cells, which were obtained from ATCC® CCL-2.1™.

2.1.3. Plasmids

Table 2: Plasmids used in this work.

Vector Name	origin	backbone	insert
pEX-A128-T7oligoctrR0332	Eurofins	pEX-A128	T7 aptamer ctrR0332
pEX-A128-T7oligoIhtA	Eurofins	pEX-A128	T7 aptamer IhtA
pEX-A128-T7oligo5s	Eurofins	pEX-A128	T7 aptamer 5s
TopoT7oligotmRNA	This work	pCR2.1-TOPO	T7 aptamer tmRNA
TopoT7oligoRNaseP	This work	pCR2.1-TOPO	T7 aptamer RNaseP

2.1.4. Oligonucleotides

The following tables list oligonucleotides used for the construction of plasmids and DNA templates for *in vitro* transcription (Table 3), oligonucleotides for sequencing (Table 4) and probes for northern blots (Table 5).

Table 3: Oligonucleotides for DNA amplification.

Name	Sequence	length (nt)	comment
MK-51	gttttttttaatacgactcactataGGGAGACCTAGCC TGGGGGTGTAAAGGTTTCGA	58	t7olgioTMRNA fw
MK-52	CTATGGAGGTGGAGAGAGT	19	t7olgioTMRNA rev
MK-53	gttttttttaatacgactcactataGGGAGACCTAGCC TCGGAAGAGTAAGGCAACCG	58	t7oligoRNASEP fw
MK-54	AGCTCGGAAGAGCGAGTAA	19	t7oligoRNASEP rev
MK-041	gttttttttaatacgactcactataGGGAGACCTAGCC TTCAAATAAAAAACTAATAAGTGGG	63	T7-oligoaptamer-ctrR0332 fw
MK-042	gttttttttaatacgactcactataGGGAGACCTAGCC TAAGTTGGTATTCTAACGCCATG	61	T7-oligoaptamer-ihtafw
MK-028	AAAGCCAAGAGAACCGGAGA	20	ihta rev
MK-024	AAAAAAAACGCAAGGCTCTGG	21	Reverse primer für sRNA ctrr0332
MK-049	GTTTTTTTTAATACGACTCACTATAGGGAGAC CTAGCCTcttggtgataatagagagagg	61	t7oligoCTR5s fw
MK-050	atgcttggcgacgacctac	20	t7oligoCTR5s rev

Table 4: Oligonucleotides used for sequencing of the plasmids.

Name	Sequence	Length (nt)	comment
MK-043	caagcccgtcagggcgcgtc	20	Eurofins A128 fw sequencing
MK-044	caggctttacactttatgct	20	Eurofins A128 rev sequencing
m13 fw	GTAAAACGACGGCCAG		universal m13 sequencing primer forward
m13 rev	CAGGAAACAGCTATGAC		universal m13 sequencing primer reverse

Table 5: Oligonucleotides used for northern blots.

Name	Sequence	length (nt)	comment
MK-014	GACCAATATATACACCCAGGCTCC	24	CtrR0332 northern blot probe front region
MK-016	GAGTCAGAAGCTATTCCATGGCGT	24	IhtA northern blot probe
MK-035	gtgcgttcgaagtgtcgatg	20	Human 5.8S northern blot probe
MK-036	caggcggtctcccatccaag	20	Human 5S northern blot probe
MK-037	atactctcgtgtatagtacc	20	*C. trachomatis* L2/434/Bu 5S northern blot probe
MK-038	CCTTAGGGCTGCTACCTTCC	20	*C. trachomatis* L2/434/Bu signal recognition particle northern blot probe
MK-039	GAGGCCAGCCAACGCCCTCC	20	*C. trachomatis* L2/434/Bu transfer-messenger- RNA northern blot probe
MK-040	CGGACTTTCCTCTGGTACTC	20	*C. trachomatis* L2/434/Bu RNAseP northern blot probe

2.1.5. Antibodies

Table 6: Primary antibodies used for western blots.

Antibody	Origin	Dilution	Company
Actin	mouse monoclonal	1:3,000	Sigma Aldrich
HSP60	mouse monoclonal	1:1,000	Santa Cruz Biotech.
Ab ββ'	rabbit polyclonal	1:2,000	Ming Tan lab
Abσ^{28}	rabbit polyclonal	1:5,000	Ming Tan lab
AB1163-4	rabbit polyclonal	1:5,000	Ming Tan lab

Table 7: Secondary antibodies used for western blots.

Antibody	Origin	Dilution	Company
goat anti-mouse IgG-HRP	goat	1:2500	Santa Cruz Biotech
goat anti-rabbit IgG-HRP	goat	1:2500	Santa Cruz Biotech

2.1.6. Molecular method Kits

Table 8: Commercial Kits.

Kit	Manufacturer
GeneJET™ Gel Extraction Kit	Thermo Scientific™
NucleoBond® PC 100	Macherey-Nagel
NucleoSpin® Plasmid	Macherey-Nagel
QIAamp DNA Mini and Blood Mini	Qiagen
miRNeasy Mini Kit	Qiagen
ERCC RNA Spike-In Mix	Invitrogen™
MEGAscript™ T7 Transcription Kit	Invitrogen™
Proteomics Dynamic Range Standard Set	Sigma-Aldrich
TURBO DNA-free™ Kit	Invitrogen™

2.1.7. Markers

Table 9: Markers used for Agarose DNA gels, Polyacrylamide RNA gels and northern blots.

Marker	Manufacturer
GeneRuler™ 1 kb DNA ladder	Thermo Scientific™
GeneRuler™ 50 bp DNA Ladder	Thermo Scientific™
PageRulerTM Prestained Protein ladder	Thermo Scientific™
RiboRuler Low Range RNA Ladder	Thermo Scientific™
pUC Mix Marker (puc57 digested with HindIII/DraI and pUC19/Msp I)	IMIB

2.1.8. Buffers, Media, and Solutions

For cell cultivation of the Hela229 cell line only RPMI 1640 from Gibco was used. Fetal calf serum (FCS) was obtained from Sigma-Aldrich. Detachment of cells was performed using TrypLE™ Express from Gibco. Washing was done using Dulbecco's Phosphate-Buffered Solution (DPBS) from GIBCO.

Table 10: Media used for bacterial cultivation.

Bacterial medium	Ingredients
LB medium	10 g tryptone 5 g yeast extract 10 g NaCl ad 1l dH_2O
LB medium (plate)	10 g tryptone 5 g yeast extract 10 g NaCl 15 g agar ad 1l dH_2O
SOC	2% (w/v) bacto-tryptone 0.5% (w/v) yeast extract 10 mM NaCl 2.5 mM KCl 10 mM $MgCl_2$ 10 mM $MgSO_4$ 20 mM glucose

SPG buffer	75 g sucrose 0.52 g KH_2PO_4 1.22 g Na_2HPO_4 0.72 g L-glutamic acid ad 1l dH_2O adjust to pH 7.4 and sterile filter

Table 11: Recipes of used buffers and solutions in this work.

Buffer	Ingredients
50x TAE	242 g Tris 57.1 ml acetic acid 37.2 g EDTA ad 1 l dH_2O
SDS upper buffer	0.5 M Tris HCl pH 6.8 0.04% (w/v) SDS
SDS lower buffer	1.5 M Tris HCl pH 8.8 0.04% (w/v) SDS
12% SDS lower gel solution (10 ml)	2.5 ml SDS lower buffer 4.0 ml 30% acrylamide 4.1 ml dH_2O 75 µl 10% APS 7.5 µl TEMED
Upper gel solution (10 ml)	2.5 ml SDS upper buffer 1.25 ml 30% acrylamide 6.25 ml dH_2O 100 µl 10% APS 20 µl TEMED
Laemmli buffer (2X)	100 mM Tris HCl pH 6.8 4% (w/v) SDS 20% (v/v) glycerol 1.5% (v/v) mercaptoethanol 0.02 g bromophenol blue
10x SDS-PAGE running buffer	30.3 g Tris 144.1 g glycine 10 g SDS
1x Semi Dry Transfer buffer	192 mM glycine 0.1% (w/v) SDS 25 mM Tris 20% (v/v) methanol

10x TBS-T	60.5 g Tris 87.5 g NaCl 5 ml Tween 20 adjust to pH 7.5 with HCl
Coomassie staining solution	44% methanol 11% acetic acid 0.2% (w/v) Coomassie R-250
Coomassie de-staining solution	20% methanol 7% acetic acid
Colloidal fixation solution	70 ml acetic acid 400 ml methanol ad 1 l dH_2O
Colloidal staining solution A	19 ml phosphoric acid (85%) 80 g ammonium sulfate ad 80 ml dH_2O
Colloidal staining solution B	2 g Coomassie G-250 ad 40 ml dH_2O
Colloidal Coomassie staining solution	40 ml Colloidal staining solution A 10 ml Colloidal staining solution B 100 ml methanol
Colloidal neutralization solution	6 g Tris 500 ml dH_2O Adjust to pH 6,5 with phosphoric acid
Colloidal washing solution	250 ml methanol 750 ml dH_2O
Silver stain fixer	50% ethanol 12% acetic acid 0.5 ml/l formaldehyde (37%)
Silver stain sensitizer	0.2 g/l $Na_2S_2O_3 \times 5H_2O$
Silver stain solution	2 g/l $AgNO_3$ 750 µl/l formaldehyde (37%)
Silver stain developing solution	60 g/l Na_2CO_3 4 mg/l $Na_2S_2O_3 \times 5H_2O$ 0,5 ml/l formaldehyde (37%)
Silver stain Stop solution	1% glycine

blocking solution	5% (w/v) BSA or milk powder in TBST-T
ECL solution 1	100 mM Tris HCl pH 8.6 2.5 Luminol 0.4 mM p-coumaric acid
ECL solution 2	100 mM Tris HCl pH 8.6 0.02% H_2O_2
10x TBE	108 g Tris 55 g boric acid 20 mM EDTA pH 8 ad 1 l dH_2O
20x SSC:	3 M NaCl 0.3 M sodium citrate adjust to pH 7 with HCl
Lysis buffer I	20 mM Tris-HCl pH 7.5 150 mM KCl 1 mM $MgCl_2$ 1 mM DTT 1 mM PMSF 0.2% Triton X-100 20 Units/ml DNaseI 200 Units/ml RNase-inhibitor
10% (w/v) glycerol solution	20 mM Tris-HCl pH 7.5 150 mM KCl 1 mM $MgCl_2$ 1 mM DTT 1 mM PMSF 0.2% Triton X-100 10% (w/v) glycerol
40% (w/v) glycerol solution	20 mM Tris-HCl pH 7.5 150 mM KCl 1 mM $MgCl_2$ 1 mM DTT 1 mM PMSF 0.2% Triton X-100 40% (w/v) glycerol
Lysis buffer II	50 mM Tris-HCl pH 8 150 mM KCl 1 mM $MgCl_2$ 5% glycerol
Wash buffer I	50 mM Tris-HCl pH 8 300 mM KCl 1mM $MgCl_2$

	5% glycerol
Wash buffer II	50 mM Tris-HCl pH 8 150 mM KCl 1mM $MgCl_2$ 5% glycerol 0.1% Triton X-100
RNA elution buffer	1.1 M Sodium acetate 0.1% SDS 10 nM EDTA
7M Urea 6% page	100 mL 10x TBE 420 g Urea 150 mL 40% acrylamide ad 1 l dH_2O
RNA loading dye GL-II (2x)	0.13% SDS 18 µM EDTA pH 8.0 95% formamide 0.025% (w/v) xylene cyanol 0.025% (w/v) bromphenol blue

2.1.9. Enzymes

Table 12: Enzymes used in this work.

Enzyme	Manufacturer
Taq DNA polymerase	Genaxxon Bioscience
Phusion High-Fidelity DNA polymerase	Thermo Scientific™
T4 DNA ligase	Fermentas
DNaseI	Thermo Scientific™
TURBO™ DNase	Invitrogen™

2.1.10. Chemicals

Table 13: Fine chemicals, inhibitors and Gel loading dyes.

Chemical	Manufacturer
Rotiphorese® Gel 40 (19:1)	Roth
Rotiphorese Gel 30 (37.5:1)	Roth
Albumin Fraction V (BSA)	Roth
Ammonium persulfate (APS)	Merck
Coomassie G-250	Roth
Coomassie R-250	Roth
dimethyl sulfoxide (DMSO)	Roth
Intas HD Green	Intas
Loading dye (6X)	Thermo Scientific™
RNA Gel Loading Dye (2X)	Thermo Scientific™
GlycoBlue	Invitrogen™
Diethyl pyrocarbonate (DEPC)	Sigma Aldrich
Roti®Phenol/Chloroform/Isoamylalcohol (25:24:1) (pH 4,5-5)	Roth
Tetramethylethylenediamine (TEMED)	Sigma Aldrich
RiboLock RNase Inhibitor	Thermo Scientific™
Dynabeads™ M-270 Streptavidin	Invitrogen™
Ethidium bromide solution 0,5 %	Roth
Phenylmethylsulfonylfluorid (PMSF)	Roth

Chemicals not listed in the table above were obtained from Sigma Aldrich, Roth and Merck.

2.1.11. Technical equipment

Table 14: Technical equipment used for this work.

Equipment	Manufacturer
Hera Cell 240i incubator	Thermo Scientific™
Hera Safe sterile bench	Thermo Scientific™
Megafuge 1.0R centrifuge	Heraeus
CT15RE centrifuge	Himac

PerfectBlue™ 'Semi-Dry'-Elektroblotter	Peqlab Biotechnology/VWR
PerfectBlue Dual Gel Twin PAGE chambers S	Peqlab Biotechnology/VWR
PerfectBlue™ Dual Gel Twin PAGE chambers M	Peqlab Biotechnology/VWR
PerfectBlue Dual Gel Twin PAGE chambers ExW S	Peqlab Biotechnology/VWR
DMIL light microscope	Leica
Scanjet G4010	HP
Shaker TR125	Infors HT
Thermo mixer comfort	Eppendorf
NanoDrop 1000 spectrophotometer	Peqlab Biotechnology
Chemiluminescence camera system	Intas
Balance ABS-80-4	Kern & Sohn
Balance EW 1500-2M	Kern & Sohn
Thermal cycler GS1	G-STORM
pH Electrode SenTix	WTW
Eppendorf 5415R - Mircocentrifuge	Eppendorf
Gradient station IP	BioComp
Optima L-80 XP Ultra Centrifuge	Beckman Coulter
Bioruptor®	Diagenode
NanoDrop™ 2000 spectrophotometer	Thermo Scientific™
Typhoon FLA 7000	GE Healthcare Life Sciences
ImageScanner III	GE Healthcare Life Sciences
dark hood DH40/50	Biostep® GmbH
Avanti J-25T centrifuge	Beckman Coulter
Avanti J-25 XP centrifuge	Beckman Coulter
Phospho storage plates	FujiFilm
FastPrep homogenizer	MP Biomedicals
GM-15 gradient maker	VWR

2.1.12. Software

Table 15: Software used for bioinformatical analysis and type setting.

Software	Publisher
Argus x1 version 7.6.17	Biostep® GmbH
Image J	Wayne Rasband (NIH)
ND-100 V3.7.1	NanoDrop Technologies, Inc. Wilmington
Endnote X9	Clarivate Analytics (previously Thomson Reuters)
Microsoft Office 365 ProPlus	Microsoft Corporation
Mendeley 1.19.3	Mendeley Ltd.
MEME version 5.0.4	(Bailey et al., 2009)
Windows 10	Microsoft Corporation
Ubunutu 14.04	Canonical Foundation, Ubuntu community
READemption	(Förstner et al., 2014)
segemehl	(Hoffmann et al., 2009)
R	R Foundation for Statistical Computing
BBMAP	(Bushnell, 2014)
python	Python Software Foundation
cutadapt	(Martin, 2011)
trimmomatic	(Bolger et al., 2014)
FastQC	(Andrews, 2010)
CorelDraw Graphics Suite X8	Corel Corporation
Integrated Genome Browser (IGB)	(Freese et al., 2016)
ImageScanner III Labscan™ 6.0	GE Healthcare Life Sciences

2.2. Methods

2.2.1. Bacterial culture methods

2.2.1.1. Cultivation of E. coli

E. coli were grown at 37 °C on LB agar plates or in LB medium shaken at 190 revolutions per minute (rpm) overnight. If selection was required, the appropriate antibiotic was supplemented into the plates or into the medium.

2.2.1.2. Stock preparation

E. coli stocks were generated by adding 150 µl Glycerol to 750 µl of *E. coli* overnight culture.

2.2.1.3. Preparation of chemo-competent *E. coli* DH5α

An overnight culture of *E. coli* DH5α was diluted by 1:100 in 100 ml LB-Medium and incubated at 37 °C at 190 rpm until an OD600 of 0.5 was reached. The bacteria were pelleted at 4000 rpm at 4 °C. The bacterial pellet was resuspended in 20 ml 0.1 M $CaCl_2$. The resulting suspension was incubated for 30 minutes (min) on ice and centrifuged again. The pellet was resuspended in 10 ml 0.1 M $CaCl_2$-solution with 20% glycerol. The bacterial suspension was then aliquoted and stored at -80 °C until use.

2.2.1.4. Transformation of chemo-competent *E. coli* DH5α

Chemo-competent *E. coli* DH5α were thawed on ice before adding the DNA. Thirty minutes after incubation on ice with the DNA, a heat shock was performed at 42 °C for 90 s followed by 2 min of incubation on ice. 800 µl of SOC medium were supplemented to the transformation and incubated for 1 hr shaking at 190 rpm and 37 °C. After incubation, the transformants was plated onto LB agar plates with the antibiotic required for selection.

2.2.2. Cell culture methods

2.2.2.1. Cultivation of Hela229 cell line

Cells were grown in 75 cm^2 cell culture flasks in RPMI 1640 medium supplemented with 10% FCS at 37 °C with 5% CO_2. To maintain a cell confluence of 70-80%, cells were passaged every two to three days. For passaging, the cells were washed with DPBS once and incubated with

1 ml of Trypsin at 37 °C and 5% CO_2 until the cells detached from the flask surface. Trypsin digestion was stopped by adding fresh 10% FCS supplemented cell culture medium. The cells were split into fresh flasks or 150 cm^2 cell culture dishes for infection with *Chlamydia*.

2.2.2.2. Preparation of cell cryo stocks

To prepare cell culture stocks, cells at a confluence of 80% were detached as described above. Five millilitres cell culture medium were added, the cells were transferred into a 15 ml tube and pelleted at 4 °C for 5 mins. Afterwards, the cell pellet was resuspended in 5 ml of 10% DMSO in FCS before cooling in 1 ml aliquots to -80 °C. For long-term storage, the cells were frozen in liquid nitrogen.

2.2.2.3. *C. trachomatis* infection

Cells at a confluency of 70% were infected with *C. trachomatis* L2/434/bu by adding the desired volume of *C. trachomatis* stock to reach the targeted multiplicity of infection (MOI) of 1. The infection was performed in medium supplemented with 5% FCS at 35 °C and at 5% CO_2 concentration.

2.2.2.4. Preparation of *C. trachomatis* stocks

Hela229 cells were grown in 150 cm^2 cell culture dishes to a confluency of 70%. The cells were infected with the appropriate volume of *C. trachomatis* for a MOI of 1. The infection was performed in s RPMI 1640 supplemented with 5% FCS. The infected cells were incubated at 35 °C and 5% CO_2 for 48 hours (hr). Using a rubber scraper, the infected cells were detached, and the infectious suspension was transferred into a 50 ml tube with glass beads. The cells were mechanically ruptured alternating vortexing and cooling on ice. After that, the supernatant was transferred into a new 50 ml tube. To remove the cell debris, the suspension was centrifuged for 10 min at 1500 x g at 4 °C. The now containing Chlamydial supernatant was centrifuged at 30,000 x g at 4 °C for 30 mins. The resulting pellet was once washed with SPG. The washed pellet was resuspended in 1 ml SPG buffer per initial 150 cm^2 cell culture dish. The Chlamydial suspension was passed through a 20G and a 18G needle, before aliquots were frozen at -80 °C for later use.

To determine the volume of Chlamydial suspension required for reaching a MOI of 1, cells with a confluency of 70% in a 12 well plate were infected with increasing volumes of a *C. trachomatis* stock. The infection of the different volumes was controlled using a bright field microscope. Thus, the volume of the suspension can be estimated to obtain the inclusion forming unit to infect every cell and to calculate the final concentration of the *C. trachomatis* stock.

2.2.2.5. Preparation of *C. trachomatis* pellets for high-throughput methods

Hela229 cells were grown in twenty cell culture dishes (150 cm^2) as described above (Preparation of *C. trachomatis* stocks, page 32) and harvested 36 hpi. Cells were scraped and ruptured as stated before but washed in SPG twice before resuspending the pellet in 2 ml of SPG and transferring into two 1.5 ml reaction tubes. The reaction tubes were centrifuged at 21500 x g for 30 min at 4 °C. The supernatant was discarded, and the pellet was snap frozen in liquid nitrogen before storage at – 80 °C.

2.2.3. DNA methods

2.2.3.1. Genomic DNA isolation

Isolation of genomic DNA from *C. trachomatis* was performed with the QIAamp DNA Mini and Blood Mini kit (QIAGEN) according to the manufacturer's manual for "Isolation of genomic DNA from biological fluids". As described in the handbook, no RNaseA digest was performed. Two 100 µl aliquots of Chlamydial stocks were used for DNA extraction. The genomic DNA was eluted in 50 µl dH_2O and concentration was measured via UV-VIS measurement.

2.2.3.2. Plasmid isolation

Plasmid isolation from *E. coli* was done either using NucleoSpin® Plasmid for miniprep plasmid isolation or NucleoBond® PC 100 for midiprep plasmid isolation. 5 ml of an *E. coli* overnight culture was used for minipreps and 50 ml of overnight cultures was used for midipreps. Both procedures were preformed according to the manufacturer´s instructions. The concentration of the isolated DNA was measured via UV-VIS measurement.

2.2.3.3. Polymerase chain reaction

Polymerase chain reactions (PCR) for amplification of DNA sequences were performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific™) in 50 µl reactions mixes:

2.5 µl	forward primer (10 µM)
2.5 µl	reverse primer (10 µM)
10 µl	5x HF buffer
0.75 µl	DMSO
1 µl	dNTPs
0.25 µl	Phusion polymerase
50 ng	vector/template
	ad 50 µl H_20

To obtain higher yields, PCR cycles were adjusted accordingly:

98 °C	5 min	
98 °C	30 s	repeat 39 times
58 °C	30 s	
72 °C	15 s (30 s/kb)	
72 °C	5 min	
4 °C	holding	

The resulting PCR products were separated via agarose gel electrophoresis and visualised under UV light.

2.2.3.4. Colony PCR

Colony PCRs were used to screen for correct plasmid insertion in transformed *E. coli*. Therefore, material from single *E. coli* colonies were picked, suspended in 20 µl dH_20 and boiled for 10 min at 94 °C. Template DNA for the reaction was crudely separated by centrifugation of the boiled *E. coli* suspension at 15,000 x g for 2 min. The supernatant was used as template DNA in PCR reactions composed as follows:

0.5 µl	forward primer (10 µM)
0.5 µl	reverse primer (10 µM)
2.5 µl	10X MolTaq buffer
0.5 µl	dNTPs
0.25 µl	MolTaq polymerase
2.5 µl	template
	ad 25 µl H_20

Colony PCR was performed according to following program:

94 °C	5 min	
94 °C	30 s	repeat 25 times
50-60 °C	30 s	
72 °C	30 s (1 min/kb)	
72 °C	5 min	
4 °C	holding	

The resulting products were separated via agarose gel electrophoresis and visualised under UV light.

2.2.3.5. Agarose gel electrophoresis

DNA was separated by size via agarose gel electrophoresis. Depending on the size of the expected fragments, PCR products were loaded onto a 1 - 2% agarose gel (with TAE buffer) containing Intas HD Green (Intas) and a voltage of 130 V was applied. The separated DNA was visualized using UV light.

2.2.3.6. Gel extraction and PCR purification

DNA fragments from agarose gels were excised and purified with the GeneJET™ Gel Extraction Kit (Thermo Scientific™) according to manufacturer´s instructions. Direct PCR product purification was performed by adding equal amounts of binding buffer to the reaction samples followed by proceeding with instructions from the Handbook.

2.2.3.7. TOPO blunt-end ligation

To construct the desired TOPO vector with the insert of interest, only TOPO blunt-end ligation was used. Therefore, PCR products generated with the Phusion polymerase had to be T/A-tailed. After the initial PCR, the products were completely frozen and thawed. 1 µl of Moltaq was added for 10 min at 72 °C before purification with the GeneJET™ Gel Extraction Kit (Thermo Scientific™). The ligation was performed according to the manufacturer´s instructions, but with half of the amount of TOPO vector. Transformation was done with chemo-competent *E. coli* DH5α.

2.2.4. Protein biochemical methods

2.2.4.1. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was performed to separate proteins by mass. Therefore, protein samples were boiled at 94 °C in 2x Laemmli buffer for 5 min. Proteins were separated under the denaturing conditions of a 12% polyacrylamide gel containing SDS. SDS masks the native charge of proteins negatively and causes their denaturation.

2.2.4.2. Coomassie staining of proteins

Visualisation of proteins separated by SDS-PAGE was performed by staining the polyacrylamide gel with Coomassie R-250. The gel was incubated for 30 min in Coomassie staining solution and de-stained by washing the Coomassie de-staining solution until bands on the gel were visible. Eventually, gels were scanned with ImageScanner III.

2.2.4.3. Colloidal Coomassie staining Proteins

Visualisation of proteins with lower concentration was performed by Colloidal Coomassie staining using Coomassie G-250. Therefore, the gel was first fixed in Colloidal fixation solution for 1 hr. After fixation the gel was incubated overnight in Colloidal Coomassie staining solution (by combining Colloidal staining solution A, Colloidal staining solution B and methanol). The gel was then incubated for 5 min in Colloidal neutralization solution and washed twice for 10 min with Colloidal washing solution. De-staining was performed by washing the gel repeatedly with dH_20 overnight. The next day, gels were scanned with ImageScanner III.

2.2.4.4. Silver Staining

Proteins were visualized by silver staining of the protein gel. The gels were fixed with silver stain fixer for 1 hr at room temperature. Gels, which were previously Colloidal Coomassie stained, were not fixed again. After fixation, the gel was washed twice for 20 min in 50% ethanol, followed by an incubation in Silver stain sensitizer. The gel was than washed with dH_20 three times and incubated with Silver stain solution, followed by two further washing steps in dH_20. After the final wash step, the gel was incubated with the Silver stain developing solution until the protein bands became visible. Gels were placed into the silver stop solution to terminate staining and scanned with ImageScanner III.

2.2.4.5. Western blot

The Detection of specific proteins from an SDS-PAGE was performed by transferring the proteins within the gel to a PVDF membrane. The membrane was activated by incubation in methanol for 1 min. The membrane was then transferred into Semi Dry Transfer buffer. Whatman paper were soaked in Semi Dry Transfer buffer. Starting from the cathode, a transfer stack was assembled (Whatman paper, PVDF membrane polyacrylamide gel, Whatman paper). Proteins were transferred at 1 mA/cm^2 membrane surface area for 2 hrs.
After protein transfer, the proteins were detected by immunoblotting. Therefore, the membrane was first incubated in BSA blocking solution. The membrane was then incubated overnight at 4 °C with the primary antibody diluted in BSA blocking solution according to Table 6. The membrane was washed thrice in TBS-T for 10 min each and incubated with a secondary antibody coupled with horseradish-peroxidase for 1 hr at room temperature. The membrane was again washed thrice for 10 min in TBS-T. After that, ECL mix (ECL solution 1 and ECL solution 2 in a 1:1 ratio) was spread evenly across the surface of the membrane. The protein of interest was than observed via a Chemiluminescence camera system.

2.2.5. Gradient methods

2.2.5.1. Preparation of glycerol gradients by Gradient Master Station

A thin-walled polypropylene Tube (14 ml, 14 x 95 mm) from Beckman coulter was filled with 10% (w/v) glycerol solution 2 mm above the centreline of the tube. 40% (w/v) glycerol solution was injected below the 10% solution until the interphase reached the centreline. After

levelling the magnetic platform of a BioComp Gradient Master Station, the tubes containing the gradient were placed into the metal rack from the device and onto the magnetic platform. To create a 10-40% linear glycerol gradient the program “10-40% (w/v) glycerol short cap” was used. The gradients were generated shortly before centrifugation and were stored at 4°C until use.

2.2.5.2. Preparation of glycerol gradients by Gradient maker device

To generate a linear gradient, a GM-15 15 ml gradient mixer was used. In both chambers of the mixing Gradient maker magnetic steers were used during pouring. The unit itself was attached to a peristaltic pump. The reservoir chamber contained the heavy 40% glycerol solution and the mixing chamber contained the 10% (w/v) glycerol solution. During pouring, the mixing valve and the outlet valve had to be opened simultaneously. The outlet from the peristaltic pump was raised short above the surface of the poured gradient in a thin-walled polypropylene tube for ultracentrifugation.

2.2.5.3. Gradient profiling (Grad-seq)

A purified Chlamydial cell pellet from forty 150 cm^2 cell culture dishes was resuspended in 500 µl lysis buffer followed by adding 750 µl of 0.1 mm glass beads. Lysis of the cells was performed by repeated cycles of vortexing for 15 s and cooling for 15 s on ice. Afterwards, the lysate was cleared by centrifugation at 13,000 rpm at 4 °C for 15 min. 10 µl of the lysate were removed for RNA extraction in 1 ml Trizol, further 20 µl of the lysate were supplemented with 20 µl 5x Laemmli buffer as controls. 200 µl of a linear 10-40% (w/v) glycerol gradient was replaced by the rest of the cleared lysate. The gradients were then centrifuged at 100,000 x g for 17 hrs at 4 °C. Fractionation of the gradient was performed manually separating it into 21 Fractions in 590 µl steps. The last fraction contained the pellet. As a control, the absorption was measured at 260 nm. Peaks for the major absorbance zones were exposed for the top region and 30S and 50S ribosomal subunits should be visible. From each fraction, except the fraction containing the pellet, 90 µl supplemented were supplemented with 5x Laemmli buffer and used for protein analysis. The 20 µl from the pellet were mixed with 20 µl 5x Laemmli buffer. The remaining Volume of each fraction underwent RNA isolation via P/C/I extraction and additional DNase digestion. The final RNA concentration quantified on a NanoDrop 2000 via UV-VIS measurement. The RNA was used for EtBr gels and northern blots and ultimately

RNA sequencing. Protein loaded onto 12% PAGE Gels, which were Colloidal Coomassie or silver stained and additionally undergone mass spectrometry (MS).

2.2.5.4. RNA isolation for the Gradient Screen

RNA extraction for the fractions was performed as follows: to each fraction 50 µl 10% SDS was added and to the pellet 25 µl P/C/I was added. The SDS supplemented fractions were mixed for 20 s followed by addition of 600 µl of acidic P/C/I. The pellet was incubated with 300 µl of P/C/I. The samples were vortexed for 30 s and incubated for 5 min at room temperature. The lysate control was mixed with 1 ml of Trizol. Four hundred microliters of chloroform were added, and the lysate control was shaken by hand for 10 s. After incubation, every sample including the control was centrifuged for 15 min at 13,000 rpm and 4 °C. The aqueous phase was collected and supplemented with 1 µl of GlycoBlue as well as with 1.4 ml of an ice-cold mixture of 100% ethanol and 3 M sodium acetate in a ratio of 30:1. Precipitation was performed at -20 °C for 1 hr followed by centrifugation at 13000 rpm for 30 min at 4 °C. The resulting RNA pellet was washed with 350 µl of 70% ethanol and air dried. The cleaned RNA was eluted in 40 µl of DEPC-treated water.

2.2.5.5. DNA digest for the Gradient Screen

The RNA samples were subsequently DNA digested. Therefore, a DNase master mix was prepared:

115 µl DNase I buffer with $MgCl_2$
11.5 µl RNase-inhibitor
92 µl DNase I
11.5 µl DEPC-treated water

The RNA was denatured for 5 min at 65 °C. After denaturing, 10 µl of the master mix was added to 40 µl of purified RNA and incubated for 45 min at 37 °C. The volume was increased by 150 µl DEPC-H_2O. Two hundred microlitres acidic P/C/I per fraction was used for a subsequent RNA extraction as described before, with the exception that for the precipitation 600 µl of 100% ethanol:3M NaOAc (30:1) was used and the RNA was eluted in 35 µl DEPC-H_2O.

2.2.5.6. Preparation of RNA and protein samples for high-throughput analysis

RNA for high-throughput sequencing was sent to vertis Biotechnolgie AG. Three microlitre of RNA from a fraction was mixed with 10 µl of 1:100 diluted ERCC (External RNA Controls Consortium) spike-in, which was used as a reference in further analysis.

Samples for mass spectrometry were prepared by mixing 20 µl from a fraction with 10 µl of UPS2 (Universal Proteomics Standard) mix. The protein samples were then homogenised using a Bioruptor set to "30 s on/30 s off high power". Samples were centrifuged at 13,200 rpm for 15 min at 4 °C and the supernatant was used for MS analysis. Both standards (ERCC and UPS2) were used in the downstream analysis as external references and for data correction.

2.2.5.7. RNA-seq analysis of the gradient

The RNA-seq data from gradient were trimmed with cutadapt (Martin, 2011) and checked with FastqQC (Andrews, 2010). The trimmed short reads were processed with READemption (Förstner et al., 2014), which uses the segemehl mapper (Hoffmann et al., 2009). The reads were mapped against the human genome, the chlamydial genome, the known ncRNAs, the annotation from the TagRNA-seq and the ERCC spike-in sequences.

Correction for sequencing differences was performed by calculating the size factors for each fraction using ERCC spike-in as a standard in a deseq2-like approach. The data were normalised per gene for its maximum occurrence using Gratitude (Di Giorgio Silvia, 2017). The data were further analysed and clustered using a self-written script (Grad-seq analysis script, p. 135). Final statistical analysis and visualization was performed in R.

2.2.6. RNA Methods

2.2.6.1. Polyacrylamide gel Electrophoresis for RNA analysis

Separation of individual RNAs was done by polyacrylamide gel electrophoresis. Therefore, 40% polyacrylamide was diluted and supplemented with 7 M Urea. This solution was polymerised by adding 0.01% APS and 0.001% TEMED. RNA samples were prepared with 2 x GLII loading buffer, boiled for 5 min at 95 °C and then loaded onto the gel. RNA was separated at 300 V for 1 hr and 50 min at RT. To visualize the RNA, the gel was incubated for 15 min in ethidium bromide solution before scanning the gel on a GE Typhoon scanner.

2.2.6.2. Northern blot

Denatured RNA (~5 µg) in GLII buffer was loaded onto a 7M Urea 6% PAGE and transferred to Hybond-XL membranes (GE Healthcare) by electro-blotting for 1 h at 50 V and 4 °C. The membrane was hybridized at 42 °C with gene-specific [32P] end-labelled oligonucleotide probe in ULTRAhyb™ Ultrasensitive Hybridization Buffer. The membrane was then exposed to a phosphor storage screen depending on signal strength and scanned on a GE Typhoon scanner.

2.2.6.3. *In vitro* transcription

RNAs for the RNA-aptamer-pull-down were generated using purified PCR products generated for the constructed vectors. The *in vitro* transcription was performed with the MEGAscript™ T7 Transcription Kit (Invitrogen™) using 500 ng of template per reaction as input. The reaction was performed overnight. The DNA of the reaction was digested with Turbo DNase for 15 mins. RNA was then purified from a 7 M Urea 6% PAGE. The gel was stained with ethidium bromide and visualized using a UV light source. The RNA bands of interest were excised. The gel slices were transferred into 1.5 ml reaction tubes filled with 750 µl of RNA elution buffer. The gel slices rotated at 4 °C overnight. The supernatant was transferred into a new tube and 750 µl of P/C/I were added before the mixture was transferred to a Phase lock heavy tube. Separation of the aqueous phase was done by centrifugation at 15,000 x g for 15 min at 4 °C. The aqueous phase was transferred to a new tube and a mixture of ice-cold 100% ethanol and 3M NaOAc in a ratio of 30:1 was added. Additionally, 1 µl GlycoBlue was added. Precipitation of RNA was performed overnight at -20 °C. The mixture was then centrifuged at 15,000 x g for 30 min at 4 °C and the resulting pellet was washed twice with 75% ethanol. The pellet was air dried and resuspended in 30 µl nuclease free water. Measurement was done via UV-VIS spectroscopy.

2.2.6.4. RNA extraction for TagRNA-seq

RNA extraction for TagRNA-seq was performed with the miRNasy kit (QIAGEN) according to the handbook without on-column DNA digestion. Pellets from twenty 150 cm^2 infected cell culture dishes were used as initial material. DNA was digested with the TURBO DNA-free™ Kit (Invitrogen™). The sample was then sent to Vertis Biotechnologie AG where the rRNA

molecules were removed with Ribo-Zero Gold rRNA Removal Kit (Epidemiology) (Illumina). Preparation of cDNA for whole transcriptome sequencing including identification of TSSs and PSSs was done by Vertis Biotechnologie AG. Initially tagged RNA adapters were ligated to 5' monophosphorylated RNA, then 5' triphosphate (5'PPP) was enzymatically conversed to 5' monophosphate (5'P) and a second RNA adapter was ligated to the newly formed 5'P of primary transcripts. Then cDNA libraries for the Illumina NextSeq 500 were generated and the sequencing itself was performed.

2.2.6.5. Bioinformatical analysis of TagRNA-Seq

The TagRNA-seq data were provided in three libraries (untagged transcripts, tagged with PSS-adapter and tagged with TSS-adapter). All reads were trimmed with trimmomatic (Bolger et al., 2014) (Supplementary information
TagRNA-seq trimming, p. 115). The quality of the reads was checked before and after trimming with FastQC (Andrews, 2010). Mapping of the short reads was performed with BBMAP from the BBTOOLs toolkit (Bushnell, 2014). The parameters for BBMAP were chosen to allow only perfect alignments and to remove ambiguous alignments.
An in-house script was written to use the reported coverage per base from BBMAP to analyse the coverage and to detect processing events, TSSs and untranslated regions (UTR) (TagRNA-seq processing script, p. 115). If abrupt rises in coverage were detected by the script, the surrounding of this change was analysed in all three libraries and it was assessed what kind of processes the change in coverage caused. Therefore, the script used the reference annotation to generate annotations for the detected changes in coverage. The script reported the changes in a basic annotation (in gff3-format). Following events were reported by the script: UTR, UTSS (transcriptional start with unknown gene), TSS (transcriptional start site within a gene), PSS (processed transcript from a gene) and UPSS (processing site with unknown gene). By manually curating this annotation in company of the coverage data in IGB, many notations were changed, and another type of event was annotated: UTRP (untranslated region from Processing Event).

2.2.6.6. Protein capture via oligo aptamer RNAs

Pull-down of specific RNAs was performed by *in vitro* transcribed RNAs with a 14 bp aptamer at the 5’-end. Hundred microliters of streptavidin-coupled Dynabeads M-270 (Invitrogen™)

were washed three times with 1 ml lysis buffer II supplemented with 0.05% Tween-20. Four microliters of 3'-biotinylated 2'-O-methyl-RNA adapter (5'-AGGCUAGGUCUCCC-3') were added to the beads and incubated for 1 hr at 4 °C. The beads with the adapter were washed twice with 1 ml lysis buffer II supplemented with 0.05% Tween-20. The beads were then suspended in 1 ml lysis buffer II with 0.05% Tween-20. Five hundred microliters of the mixture were stored overnight at 4 °C for later use. For RNA capture, 10 µg of RNA were incubated with the adapter-coupled beads over night at 4 °C while rotating. For one RNA of interest the chlamydial pellets of 20/ 30 cell culture plates were resuspended in 500 µl lysis buffer II with addition of 1 mM DTT and 1 mM PMSF. The chlamydial pellets were lysed using lysis matrix B tubes in a FastPrep (MP medicals) at 5 m/s for 30 s. The lysate was cleared at 13,000 rpm for 30 min at 4 °C. If multiple samples were analysed, the different supernatants were pooled and mixed. The beads without the RNA of interest were used to clear the lysate by adding 500 µl of the lysate supernatant to the adapter-coupled beads. This mixture was rotated for 3.5 hrs at 4 °C. Fifteen microliters of the supernatant was taken as an input control and the rest of the supernatant was mixed with the RNA-coupled beads. The adapter-coupled beads, which were used for preclearing of the lysate, were discarded. RNA-coupled beads with lysate incubated for 2 hrs while rotating at 4 °C. The beads were washed with wash buffer I, wash buffer II and lysis buffer II each supplemented with 1 mM DTT and 1 mM PMSF. The lysis buffer was removed, and the beads were boiled in 2x Laemmli or 1x LDS sample buffer. Using a magnetic rack, the beads were removed. The protein samples were then used for SDS-PAGE and for MS.

2.2.7. Mass spectrometry analysis of the Grad-seq and oligo aptamer Pull-down

2.2.7.1. In-solution and in-gel digestion

MS preparation and analysis were performed by employees of AG Schlosser, Rudolf Virchow centre for experimental Biomedicine, University of Würzburg. The gradient samples were digested in-solution while the samples for the oligonucleotide aptamer pull-down were in-gel digested.

For the protein samples from the gradient, proteins were reduced with 125 mM DTT for 10 min at 70 °C. Alkylation was performed for 20 min with 270 mM iodoacetamide. The proteins were precipitated overnight at -20 °C using the fourfold volume of acetone. The resulting pellet was dissolved in 50 µl 8 M urea in 100 mM ammonium bicarbonate. The proteins were

digested with 25 µg Lys-C (Wako) for 2 hrs at 30 °C. The samples were diluted to 2 M urea by adding 150 µl of 100 mM ammonium bicarbonate followed by digestion with 0.25 µg trypsin at 37 °C overnight. C-18 Stage Tips were used for desalting (Rappsilber et al., 2003). Elution of the peptides was performed with 60% acetonitrile in 0.3% formic acid. Peptides were dried with a laboratory freeze-dryer and sored at -20 °C. Before nanoLC-MS/MS analysis, the samples were dissolved in 2% acetonitrile / 0.1% formic acid.

Proteins from the oligonucleotide aptamer pull-down were precipitated with the fourfold volume of acetone overnight at -20 °C. The resulting pellets were washed with acetone three times. The proteins were taken up in NuPAGE® LDS sample buffer (Life Technologies). The samples were reduced with 50 mM at 70 °C for 10 min, before they were reduced with 120 mM Iodoacetamide for 20 min. Separation of proteins by mass was performed with a NuPAGE® Novex® 4-12% Bis-Tris gels (Life Technologies) according to manufacturer´s instructions. Visualization of the proteins was done by first washing the gel three times for 5 min each followed by staining with Simply Blue™ Safe Stain (Life Technologies). The gel lanes were cut into 15 pieces after another wash step in water for 2 hrs.

For the in-gel digestion the gel slices were de-stained with 30% acetonitrile in 0.1 M NH_4HCO_3 (pH 8). The gel slices were shrunk with 100% acetonitrile. A vacuum concentrator was used for drying the samples. The final digestion was performed with 0.1 µg trypsin per gel band overnight at 37 °C in 0.1 M NH_4HCO_3 (pH 8). Peptides were extracted from the gel-slices with 5% formic acid after removing the supernatant. Samples were then pooled and used for nanoLC-MS/MS.

2.2.7.2. NanoLC-MS/MS Analysis

The NanoLC-MS/MS analysis was performed on an Orbitrap Fusion (Thermo Scientific) for the gradient profiling and the aptamer pull-down. This device was equipped with a PicoView Ion Source (New Objective) and coupled to a liquid chromatography system (EASY-nLC 1000, Thermo Scientific). The samples were loaded onto capillary columns (PicoFrit, 30 cm x 150 µm ID, New Objective), which were self-packed with ReproSil-Pur 120 C18-AQ, 1.9 µm (Dr. Maisch). The peptides were separated with a 30-minute linear gradient from 3% to 30% acetonitrile and 0.1% formic acid. The flow rate was 500 nl/min.

The MS and MS/MS measurements were performed on the Orbitrap Fusion (Thermo Scientific). The resolutions for the MS and MS/MS scans were 60,000 and 15,000. 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 done with a repeat count of 1 and the exclusion duration was 30 s. Single charged precursors were excluded from the selection. The minimal signal threshold for precursor selection was 50,000. Predictive AGC was used with target value of $2e^5$ for MS scans and $5e^4$ for MS/MS scans. EASY-IC was used for internal calibration.

2.2.7.3. MS data analysis

The Data acquired from the measurements were analysed with MaxQuant version 1.6.2.2 (Cox and Mann, 2008). Andromeda was used within MaxQuant to perform the database search against a UniProt *C. trachomatis* database, the UniProt Human database and a database with common contaminates. Additionally, the gradient measurements were searched against a database containing proteins from the UPS2 proteomic standard. The tryptic cleavage specificity for the search allowed 3 mis-cleavages.

Protein identification was controlled by a false-discovery rate of 1% for protein and peptide levels. Altered from the default settings of MaxQuant following modifications were allowed: Protein N-terminal acetylation, Gln to pyro-Glu formation (N-term. Gln) and oxidation (Met). Carbamidomethyl (Cys) was set as fixed modification. Proteins were quantified by LFQ intensities (Cox et al., 2014), but proteins with less than two identified razor/unique peptides were dismissed.

The results from the aptamer pull-down were further analysed with a R script developed by AG Schlosser. Median intensities were calculated to discriminate for non-specifically enriched proteins. If the LFQ intensities in the control samples were missing, values close to the baseline were imputed. Identification of boxplot outliers in intensity bins of at least 300 proteins were used to describe enriched proteins. The log2 transformed protein ratios between sample and control were considered significantly enriched if the ratios were values outside a 1.5x (potential) or 3x (extreme) interquartile range (IQR).

3. Results

3.1. Analysis of TagRNA-Seq

The reference annotation of the chlamydial genome accounts for 937 genes and the plasmid codes for 8 genes. IhtA was the only well previously studied small ncRNA.

Previous transcriptomic studies of *C. trachomatis* focused either on total RNA or on TSSs. In a previous dRNA-Seq analysis by Albrecht and colleagues 338,678 reads were mapped to the chlamydial genome and the plasmid. By characterization of transcriptional start sites (TSSs), nine new ncRNAs were identified with 8 being found on the genome and one on the plasmid (Albrecht et al., 2011, Albrecht et al., 2010).

The aim of the present work was to characterize the intergenic regions of the 1.04 Mb chlamydial genome. Therefore, TagRNA-seq of *C. trachomatis* L2/434/bu was performed to observe the different transcripts originating from TSSs and processing sites (PSSs).

After trimming and mapping, 13,270,884 reads were mapped against the chlamydial genome and the plasmid. Out of the 13,270,884 reads, 2,017,491 reads originated from TSSs and 529,679 came from processing events. Furthermore, 10,723,714 reads were not the result of the labelled RNA species and showed the total RNA distribution within this RNA sample.

The library with the unassigned reads (10,723,714 reads) had the highest number of reads and a high sequencing depth (Figure 4). In contrast, the other two libraries didn't obtain a high sequencing depth and had less reads. The TSS library had a higher read count and more regions with a high coverage compared to the PSS library.

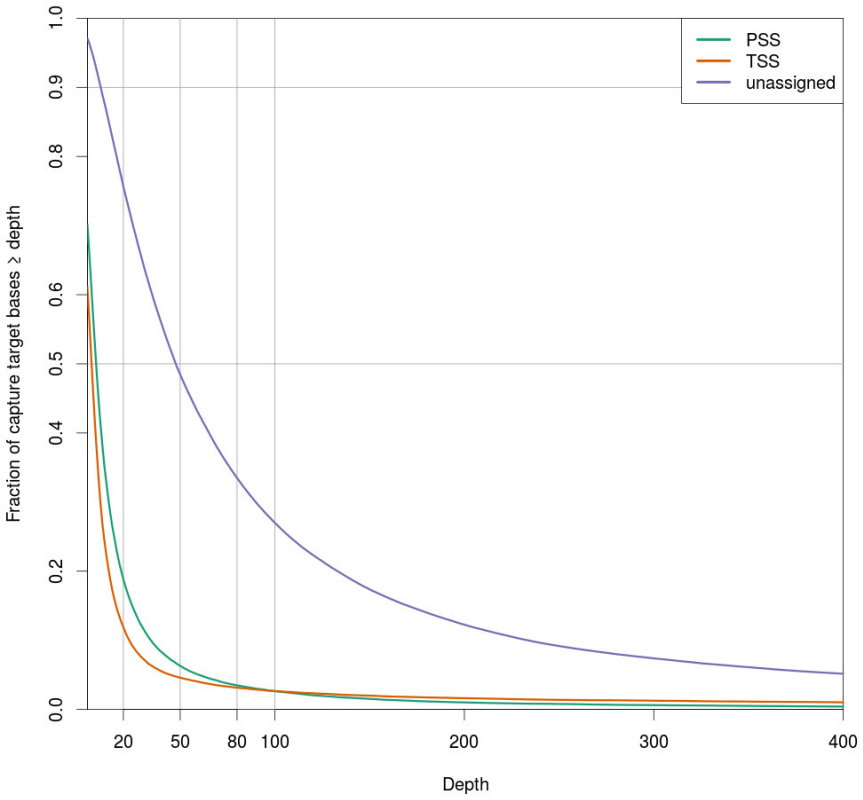


Figure 4: Coverages of the three libraries enriched for processing events (green line), transcriptional start sites (orange line) and unassigned RNAs (blue line).

After mapping, a program (TagRNA-seq processing script, p. 115) for initial analysis of the resulted coverage files from the PSS-, TSS- and total RNA libraries (TagRNA-seq processing script, p.115) was written. This program assessed if sudden rises could be associated with a TSS or a PSS. If multiple abrupt changes were in close vicinity, only the signal of the strongest intensity was used. These signals were then associated with the genes from the reference annotations (references: AM884176.1 & AM886278.1 from NCBI genomes). TSS signals within a gene resulted in a new annotation, starting with the TSS signal and ending with the original gene end. A TSS present within 100 bp upstream of an annotation, resulted in an annotation starting from the TSS signal to gene start marked as 5' untranslated region (UTR). PSS events within genes resulted in annotations broken up into slices by the processing events. Processing events in intergenic regions resulted in an annotation +/- 20 bp from the signal.

Taken together, this resulted in 2,250 annotations (Table 16). These annotations were then manually curated and annotations that overlapped with the reference annotation were removed.

Table 16: Type and number of new annotations based on the TagRNA-seq data after initial analysis using the self-developed python script.

Type of new annotation	Number of new annotations
UTR (5' untranslated region)	270
TSS (transcriptional start site within a gene)	361
UTSS (transcriptional start with unknown gene)	197
UPSS (processing site with unknown gene)	239
PSS (processed transcript from a gene)	1183

As a result, 1,746 new annotations remained. Out of these, 576 were in intergenic regions and were previously not described. The following table shows the distribution of the new annotations based on their affiliation (Table 17). Additionally, a new annotation type was created: UTRP (5' untranslated region from processing event).

Table 17: Type and number of new annotations based on the TagRNA-seq data after initial analysis with the self-developed python script and manual curation.

Type of new annotation	Number of new annotations
UTR (5' untranslated region)	290
UTRP (5' untranslated region from processing event)	57
UTSS (transcriptional start with unknown gene)	154
UPSS (processing site with unknown gene)	75
PSS (processed transcript from a gene)	935
TSS (transcriptional start site within a gene)	235

The intergenic types were analysed for their length and GC content (Table 18). The 5' untranslated region originating from PSSs were the fewest and longest but were varying the most. Unknown TSSs were on average the longest and are closest with their GC content to global GC content (41.3%). The other annotated sequences were below the global GC content.

Table 18: Sequence statistics for the intergenic sequences found by TagRNA-seq.

	Sequence length				GC-content in %			
Type	min	max	mean	median	min	max	mean	median
UTRP	5	2441	120.11 ± 353.82	40	19.05	55.56	35.25 ± 7.83	36
UTR	9	832	70.56 ± 70.81	50	11.11	61.11	36.3 ± 8.27	37.3
UPSS	41	179	48.89 ± 24.75	41	19.51	51.22	37.9 ± 6.76	39.02
UTSS	52	1056	157.7 ± 129.73	101	29.7	53.33	41.53 ± 4.51	41.78

Most of the RNAs have an average length of 157 bp, which ideally characterise them as sRNAs. The longest identified RNA with 1056 bp is located on the opposite strand of *recD,* which encodes a 5`-3` helicase and is part of the RecBCD holoenzyme, and *CTL0289*, which is a putative membrane transport/efflux protein. Due to the composition of the reads, two RNAs were annotated in this locus. A highly transcribed region which overlaps the 3'-end of *recD* and extends 35 bp over the end, was detected and makes up a shorter transcript. In addition, a transcribed region reaching into the middle of the upstream gene of *recD*, *CTL0289*, was observed and makes up the longer transcript (Figure 5).

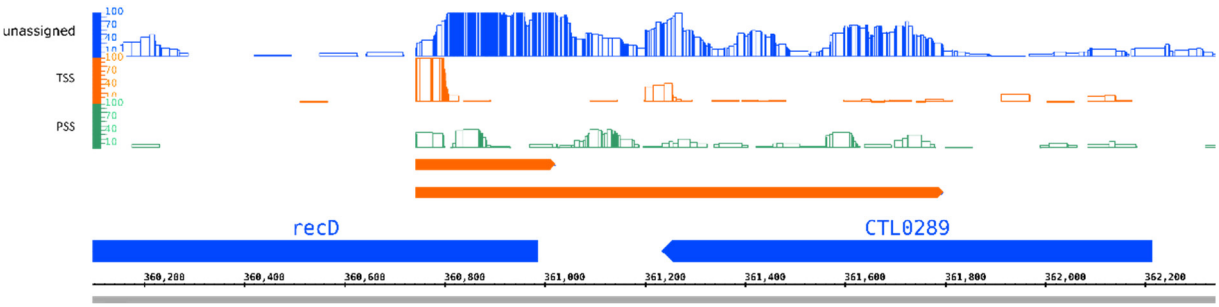


Figure 5: Visualization of the reads from the TagRNA-seq Data with the reference annotation in blue and the TSS annotation in red at the locus between *recD* and *CTL0289*. The graphs represent the mapped reads in the respective labelled library. Unlabelled reads are represented in blue, reads associated with Transcriptional start sites (TSS) are shown in red and green are reads from processing events (PSS) in the transcriptome.

The small RNA IhtA is transcribed with the downstream outer membrane ring protein *sctC*. *ihtA* possess, in addition to the TSS of the downstream gene, its own TSS. With the second TSS the abundance of the *ihtA* reads strongly increase in this locus (Figure 6).

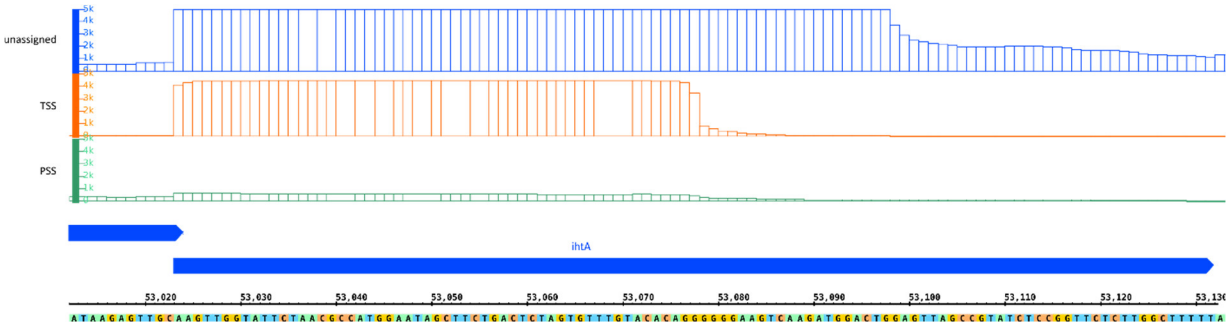


Figure 6: Visualization of the reads from the TagRNA-seq Data with the reference annotation in blue at the locus of small RNA ihtA. The graphs represent the mapped reads in the respective labelled library. Unlabelled reads are represented in blue, reads associated with Transcriptional start sites (TSS) are shown in red and green are reads from processing events (PSS) in the transcriptome.

The small RNA CtrR0332 shows to have two sites for primary transcripts and two processing sites, making this locus a highly active and transcriptional complex region. The two identified regions of the transcripts define the known processed forms of CtrR0332.

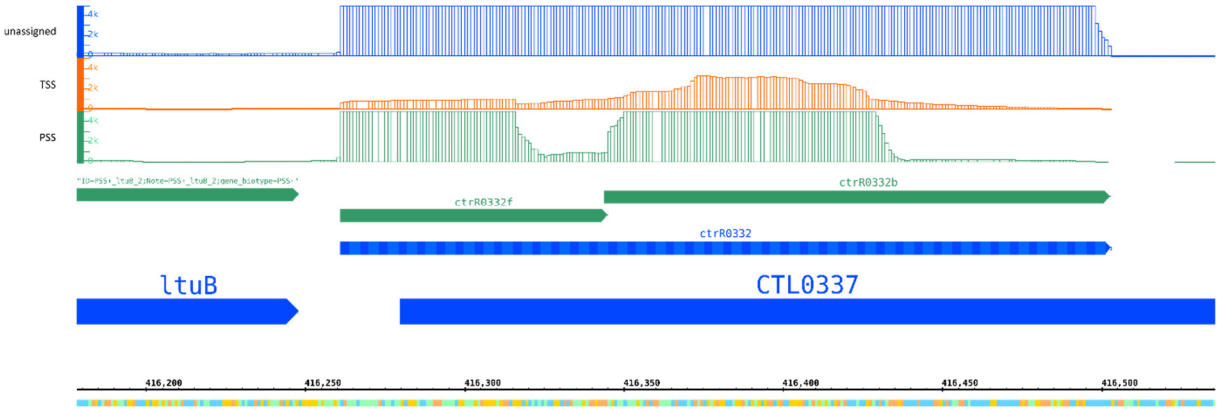


Figure 7: Visualization of the reads from the TagRNA-seq Data with the reference annotation in blue and the PSS annotation in green at the locus the small RNA ctrR0332. The graphs represent the mapped reads in the respective labelled library. Unlabelled reads are represented in blue, reads associated with Transcriptional start sites (TSS) are shown in red and green are reads from processing events (PSS) in the transcriptome.

Furthermore, the intergenic sequences were analysed with the MEME suit. Additionally, every sequence 40 bp upstream from the annotation was analysed.

Analysis of the 40 upstream sequences of the new intergenic annotations identified four motifs. The first motif (Figure 8) was found in 81 sequences from the newly annotated intergenic regions.

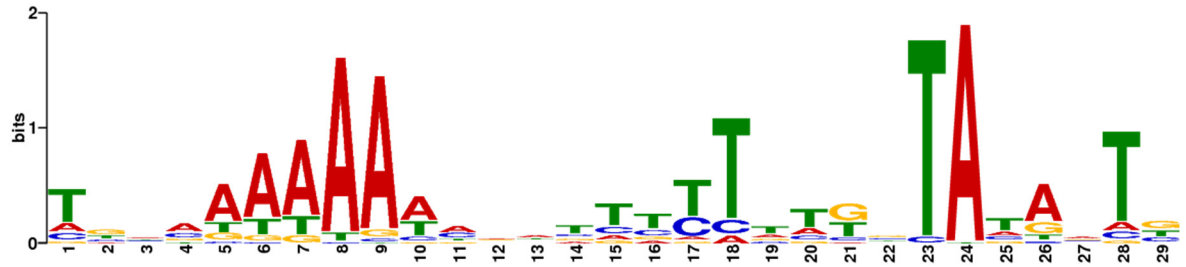


Figure 8: The first motif logo generated by the MEME suit from 40 bp upstream sequences of all new annotated intergenic sequences.

The highest association for this motif in the database is with *rpoD* (σ^{70} factor from *E. coli*). Albrecht and colleagues observed a similar sequence logo in their studies, while mapping TSSs with dRNA-seq data. The authors associated this motif to the σ^{66} promoter sequence (Albrecht et al., 2011, Albrecht et al., 2010). In addition, many motifs were associated with ferric uptake (Table 19).

Table 19: Results from tomtom from the MEME tool suit searched against all prokaryotic DNA databases for the first motif found using the 40 bp upstream sequences from the new annotated intergenic regions.

Motif (Database entry)	p-value	E-value	q-value
rpoD (rpoD17)	3.91E-04	1.93E-01	3.86E-01
DegU (MX000168)	8.26E-04	4.08E-01	4.07E-01
PhoP (MX000057)	1.40E-03	6.94E-01	4.62E-01
Fur *N. gonorrhoeae* (EXPREG_00000ec0)	2.83E-03	1.40E+00	6.93E-01
MetR (MX000158)	4.22E-03	2.08E+00	6.93E-01
CcpA *C. difficile* (EXPREG_00000d10)	6.03E-03	2.98E+00	8.50E-01
MtrB (MX000015)	9.55E-03	4.72E+00	9.99E-01
ComK (MX000023)	1.57E-02	7.74E+00	9.99E-01
ScrR (Gammaproteobacteria)	1.69E-02	8.36E+00	9.99E-01
Fur *H. pylori* (EXPREG_00000340)	1.76E-02	8.69E+00	9.99E-01
Fur *E. coli* (EXPREG_000007c0)	1.76E-02	8.70E+00	9.99E-01

The second motif, found in 77 of the sequences, was like the first one, only shifted and shortened to 21 positions, with an adenine stretch in the middle (Figure 9).

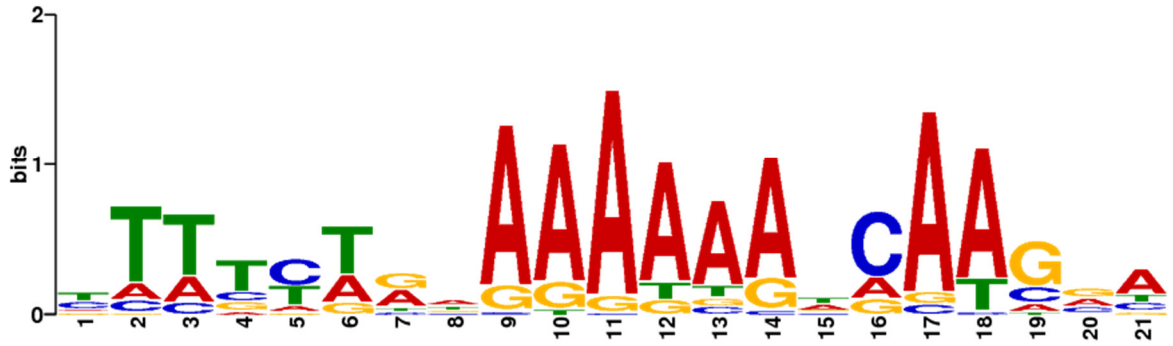


Figure 9: The second motif logo generated by the MEME suite for the 40 bp upstream sequences of all new annotated intergenic sequences.

Using tomtom, similar sequences were identified (Table 20). Many of these motifs were associated with σ-factors or response regulator binding motifs.

Table 20: Results from tomtom from the MEME tool suit searched against all prokaryotic DNA databases for the second motif found using the 40 bp upstream sequences from the new annotated intergenic regions.

Motif (database entry)	p-value	E-value	q-value
MogR *L. monocytogenes*	1.46E-05	7.20E-03	1.43E-02
CovR (MX000041)	2.97E-04	1.47E-01	9.69E-02
CodY *S. pyogenes* (EXPREG_00000330)	7.79E-04	3.85E-01	1.53E-01
NagC (MX000150)	2.42E-03	1.20E+00	3.45E-01
SigL (MX000078)	2.55E-03	1.26E+00	3.45E-01
Fnr (MX000004)	3.35E-03	1.65E+00	3.45E-01
LexA *S. meliloti* (EXPREG_00001480)	3.48E-03	1.72E+00	3.45E-01
ExsA (MX000103)	4.37E-03	2.16E+00	3.45E-01
CodY *L. lactis* (EXPREG_000001b0)	4.56E-03	2.25E+00	3.45E-01
SigB (MX000073)	4.58E-03	2.26E+00	3.45E-01
MetR (MX000158)	6.76E-03	3.34E+00	4.73E-01
Fnr (Fnr_Gammaproteobacteria)	9.72E-03	4.80E+00	5.94E-01
RofA (MX000039)	9.80E-03	4.84E+00	5.94E-01
PhoP *Y. pestis* (EXPREG_00000050)	1.03E-02	5.09E+00	5.94E-01
OmpR (MX000142)	1.30E-02	6.41E+00	6.31E-01
lrp (lrp)	1.33E-02	6.59E+00	6.31E-01
rhoD (rpoD16)	1.36E-02	6.71E+00	6.31E-01
RpoE-SigE (MX000037)	1.42E-02	6.99E+00	6.31E-01
rhoD (rpoD15)	1.58E-02	7.81E+00	6.54E-01
SigB (MX000075)	1.60E-02	7.92E+00	6.54E-01
LexA *C. crescentus* (EXPREG_000014f0)	1.82E-02	9.00E+00	6.86E-01
Zur *P. protegens* (EXPREG_00000e20)	1.89E-02	9.33E+00	6.86E-01
HexR *S. oneidensis* (EXPREG_000012b0)	1.89E-02	9.33E+00	6.86E-01

The third motif, which was found in the 40 bp upstream sequences, was present in eight sequences and had a high association with the fur (ferric uptake regulation) binding motif (Appendix table 23, Figure 10).

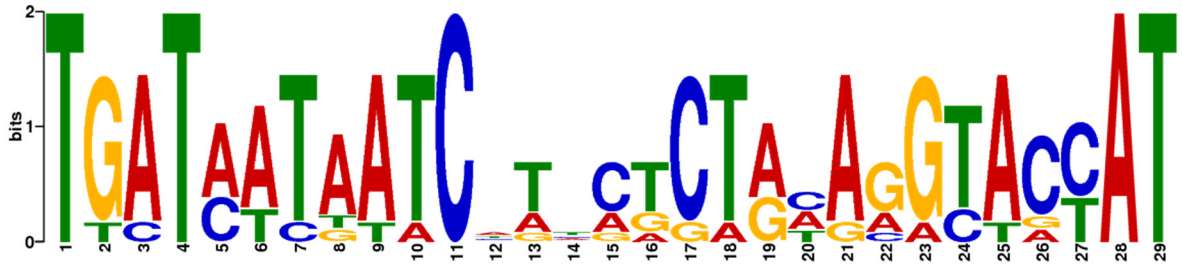


Figure 10: The third motif logo generated by the MEME suite for the 40 bp upstream sequences of all new annotated intergenic sequences.

One further identified motif was generated from 6 sequences (Figure 11) and was mostly associated with LexA-type motifs (Appendix table 24).

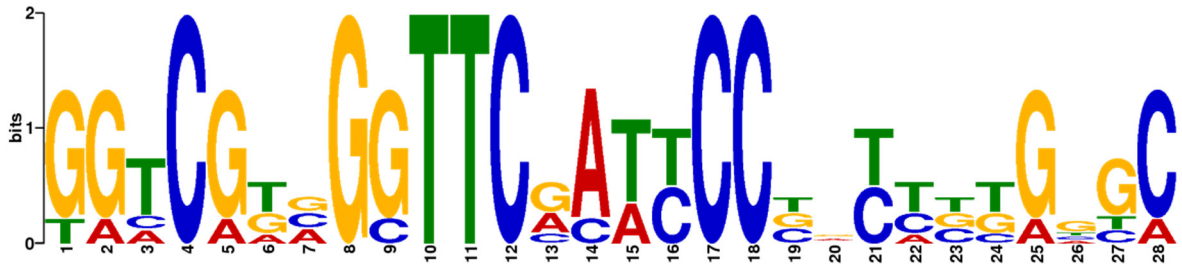


Figure 11: The last motif logo mostly associated with LexA-type motifs. The logo was generated by the MEME suite for the 40 bp upstream sequences of all new annotated intergenic sequences.

3.2. Establishment of Grad-seq

Gradient profiling by Sequencing (Grad-seq) was previously performed only in free-living bacteria (e.g. *Salmonella*) (Smirnov et al., 2016b). Therefore, the aim of this work was to establish the conditions for successful gradient profiling coupled to high-throughput methods for *C. trachomatis*. Four parameters had to be adjusted: 1) input concentration of *C. trachomatis* for the experiment; 2) the lysis of *C. trachomatis*; 3) generation of the glycerol gradient; and 4) how to inspect the resulting gradient.

The first three gradients were performed with pellets from 30 plates of cells infected with *C. trachomatis* L2/434/Bu with gradients ranging from 1% - 40% (w/v) glycerol using a gradient maker device. Only 20 fractions were taken for analysis, with the heaviest fraction containing the pellet. Only protein gels and western blots were performed for this experiment. Lysis was performed at 6.5 m/s for 20 s repeated for 5 times before the lysate was cleared.

The gradient showed a partial separation (Figure 12). Most proteins were in the lighter molecular weight fractions (1-12), as seen in the protein gel (Figure 12A). Actin only appeared in the very light fractions of the gradient (1-5), while the heat shock protein 60 spanned across 10 fractions in the middle of the gradient. Since a single protein itself would be present only in very light fractions, one can conclude that this is a complex. In *E. coli*, GroEL, which belongs to the chaperonin family, forms a tetradecamer (Braig et al., 1994) while in other organisms the paralog can function as a dimer (Qamra et al., 2004). This spread of GroEL can be explained by either fragmentation of the homo complex, by the formation of a complex of GroEL with other subunits of the folding machinery, or it is functioning as a chaperon with other proteins. The σ^{66}-factor has also been detected in these fractions. This σ-factor plays a key role in the regulation of most chlamydial proteins (Koehler et al., 1990, Lonetto et al., 1992, Paget and Helmann, 2003) and directly interacts with the β-subunit of the RNAP (Rao et al., 2009b). This interaction is indicated by the shift of the β' subunit in the western blot. The σ-factor and the RNAP subunit larger signal can be observed in the same fractions. The northern blot analysis of the small RNA ctrR0332 showed a signal present in all fractions. The strongest signal of the probe was overserved in the fractions 5 - 8. The strong signals indicate a complex for this sRNA.

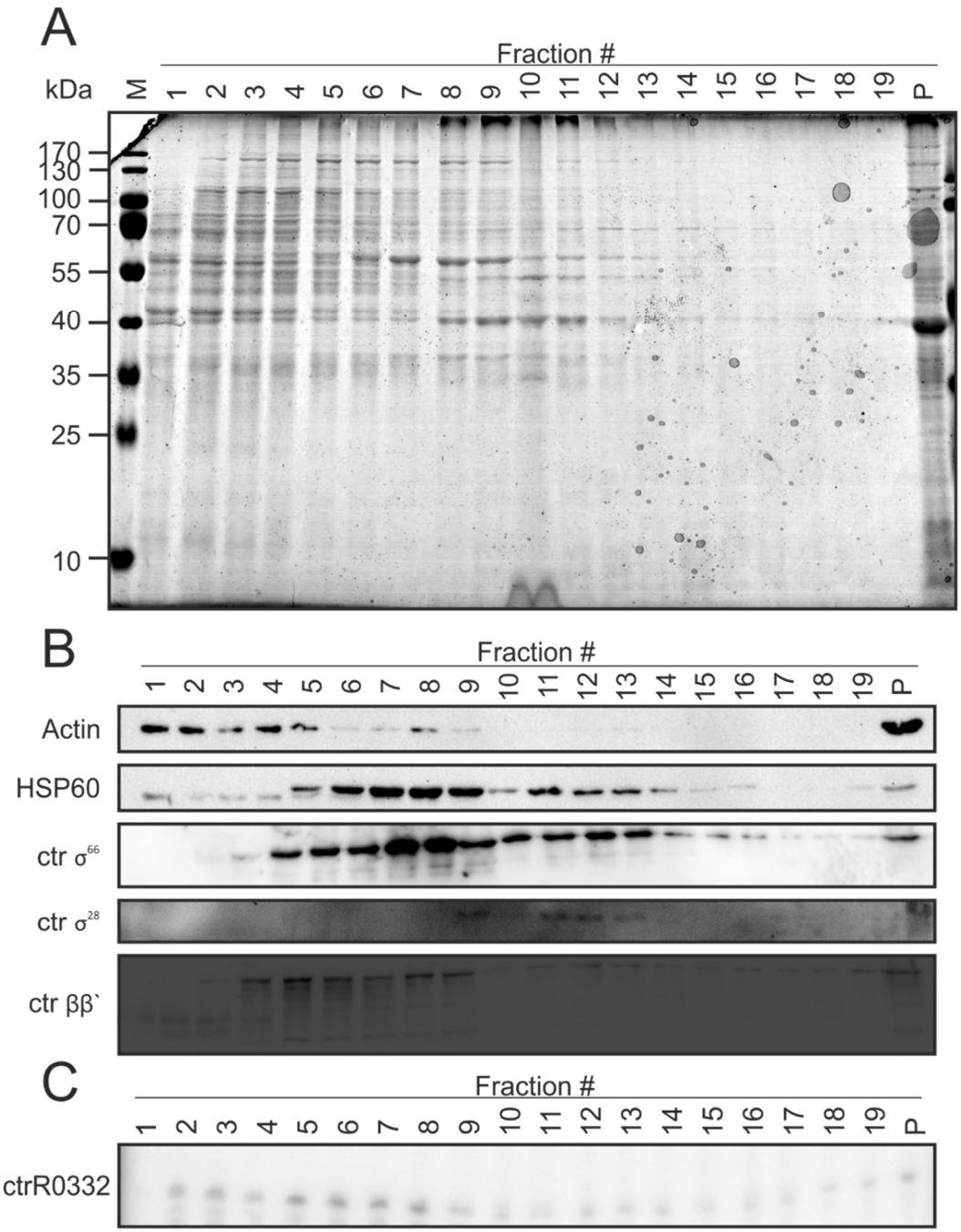


Figure 12: Pellets from thirty 150 cm^2-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep for 5 times at 6.5 m/s for 20 s and separated on a 1%-40% glycerol gradient that was made by a gradient maker device. (A) Proteins were separated using a 12% SDS-PAGE and visualised via colloidal Coomassie. The first lane shows the marker (M) and the following lanes are the fractions from the gradients collected from top to the pellet (P). (B) Proteins were separated by a 12% SDS-PAGE. Actin, the chlamydial HSP60, the chlamydial σ^{28} and σ^{66} factor and the ββ' subunit of the RNA polymerase were analysed by immunoblotting. (C) northern blot of the ctrR0332 transcript of the gradient in descending order from top of the gradient to the pellet of the gradient.

The gradient was repeated twice with 30 dishes of cells infected with *Chlamydia* but results from the western blots showed different patterns (Figure 13). Actin showed a similar pattern except for its location in the gradient. The signal was detected in fraction 7 (Figure 13A) or in fraction 8 (Figure 13B) for the other replicate. Heat shock protein (HSP60) in these gradients

shifted towards a lighter fraction starting in fraction one (fraction for the lightest complexes/molecules) and ending in the middle of the gradient. The β´-subunit of the RNAP is present only in one gradient without the lighter observable band (Figure 13A). In the other gradient (Figure 13B), degradation of the polymerase was observed, although the presence of the signals was in line with the first gradient (Figure 12B). The σ^{66}-factor was detected in each fraction of one gradient (Figure 13A) while in the other gradient, it was only present in the factions 6, 7 and 8 (Figure 13B). In these fractions the highest signal intensity was observed for one of the gradients (Figure 13A). The σ^{28}-factor was observed only in the first two gradients. In the western blot analysis of the first gradient (Figure 12B), it was present in the factions 9 to 13, while in the next gradient it was only detected in the first five fractions.

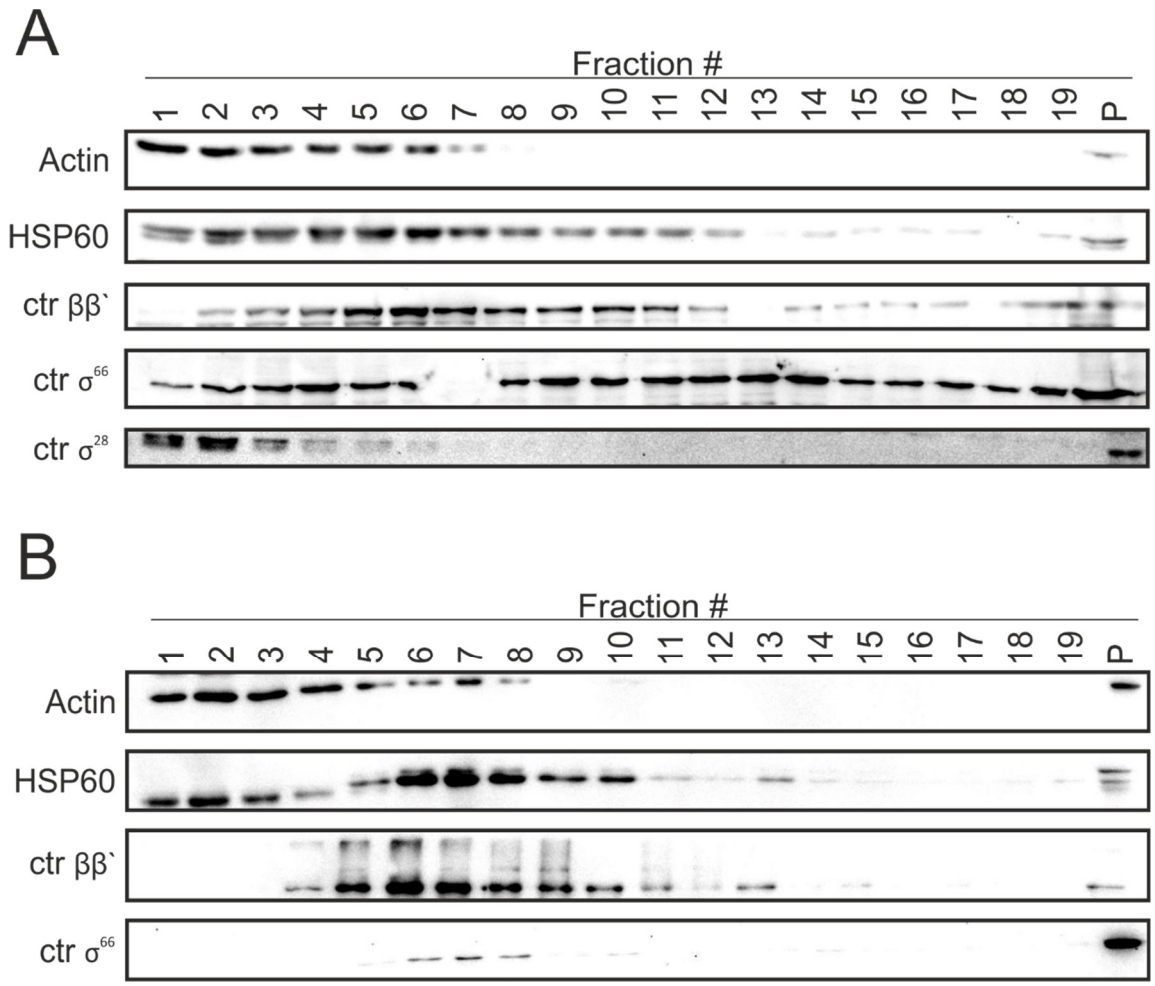


Figure 13: Pellets from thirty 150 cm^2-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep for 5 times at 6.5 m/s for 20 s and separated on a 1%-40% glycerol gradient that was made by a gradient maker device. Analysis was done via immunoblotting (A) Second replicate and (B) third replicate of the gradient with following proteins being specifically analysed: actin, chlamydial HSP60, ββ'-subunit of the RNA polymerase and chlamydial σ^{66}-factor. In the second replicate, also σ^{28}-factor was analysed.

For the next gradient, the speed of the FastPrep for the lysis was decreased from 6.5 m/s to 6 m/s repetitions, but time stayed the same as before. The gradient was similar to the first gradients. The stained gel, although similar to the first shown gel, revealed the lack of bands in the pellet lane and the distinguishable strong bands in the middle of the gradient. The western blots results were also like the previous gradients. Only the β'-subunit showed distinctly smaller bands (Figure 14).

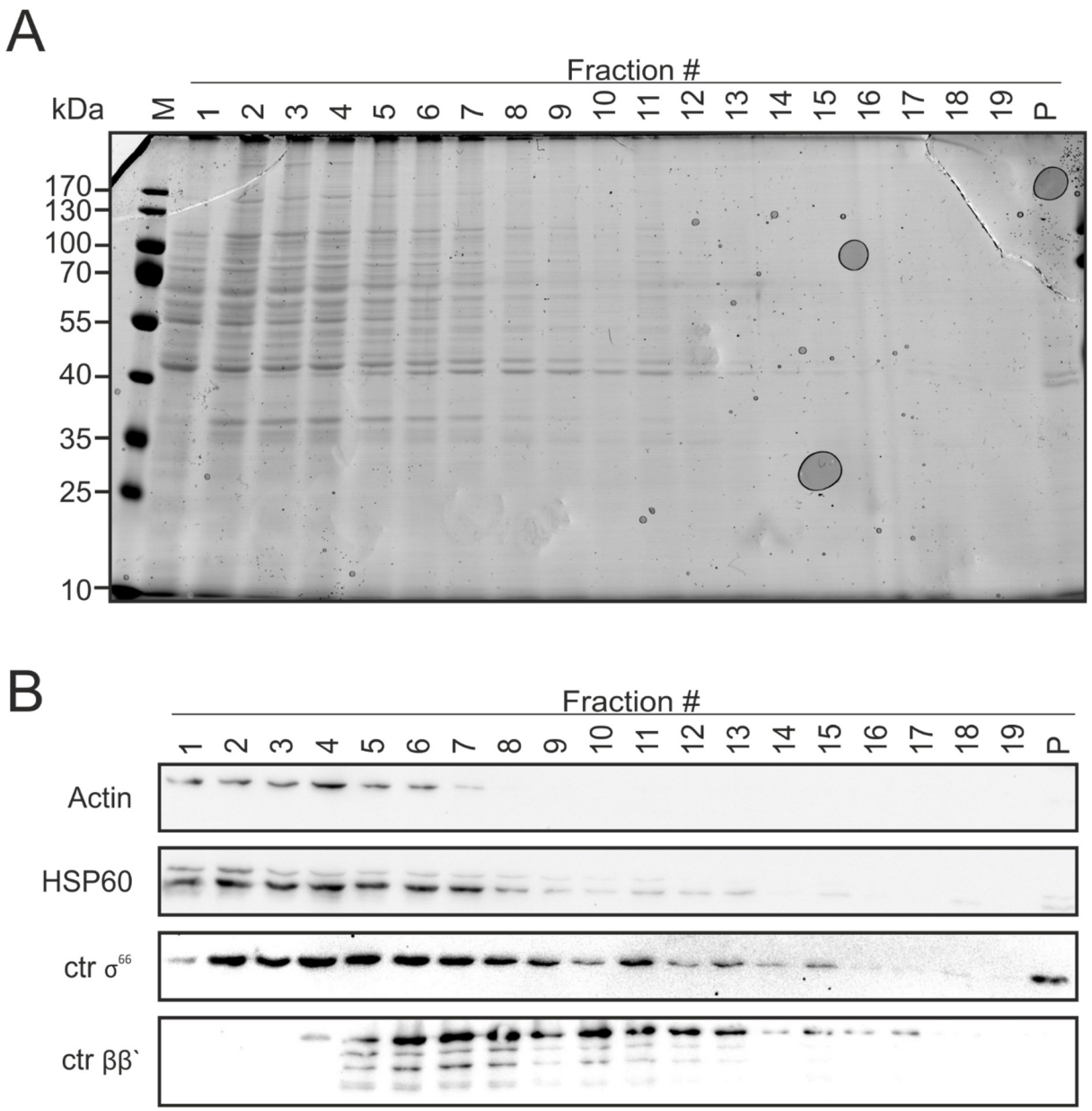


Figure 14: Pellets from thirty 150 cm^2-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep for 5 times at 6.0 m/s for 20 s and separated on a 1%-40% glycerol gradient that was made by a gradient maker device. (A) Proteins were separated using a 12% SDS-PAGE and visualised via colloidal Coomassie. The first lane shows the marker (M) and the following lanes are the fractions from the gradients collected from top to the pellet (P). (B) Proteins were separated by a 12% SDS-PAGE. Actin, the chlamydial HSP60, the chlamydial σ^{66} factor and the ββ'-subunit of the RNA polymerase were analysed by immunoblotting.

For the following gradients, instead of a gradient maker device, a gradient master station was used to automatically mix the gradients and thus get more reproducible results. Additionally, the lysis procedure was changed by removing the repetitions in the FastPrep, so that the samples were in the FastPrep only once.

In order to avoid inconsistencies from western blots, northern blots to analyse specific RNAs were performed. The subsequent gel was prepared using the above-mentioned method for gradient generation. The lysis was done at 5 m/s for 20 s. In the silver stained protein gel, possible ribosomal proteins were detected in the heavy fractions. Furthermore, in the pellet, the signal was so strong, individual protein bands were not detectable anymore, indicating an improved separation of the lysate in the gradient (Figure 15A). The RNA gel also showed a clear separation of abundant RNA species. rRNAs were seen as bands in fraction 7 to 20 and in the pellet (Figure 15B). 16S and 23S RNAs were detected below the loading pockets of the gel indicating their high molecular weights. 16S RNA was below the 23S RNA and was only detectable in the fractions 8 to 13 of the gradient. The 5S RNA was detectable only in the heavy fractions starting from fraction 13 and additionally in the pellet. The 5.8S RNA was seen above the 5S RNA but showed a less intense signal. Only in the very light fraction of the gradient (below 100 bp), tRNAs were observed (Figure 15B).

All fractions were screened for the presence of ctrR0332 and IhtA as indicators of sRNAs. IhtA was detectable in the light fractions (1 to 8), having its highest signal intensity in fraction 1, indicating that it is not present in a complex. Due to our probe for ctrR0332 the full-length sRNA and the 80 bp processed 5’ fragment, which was called ctrR0332’, were detectable. Both fragments were observed across the gradient and had their peak in the fractions 7 and 8 as well as in the pellet.

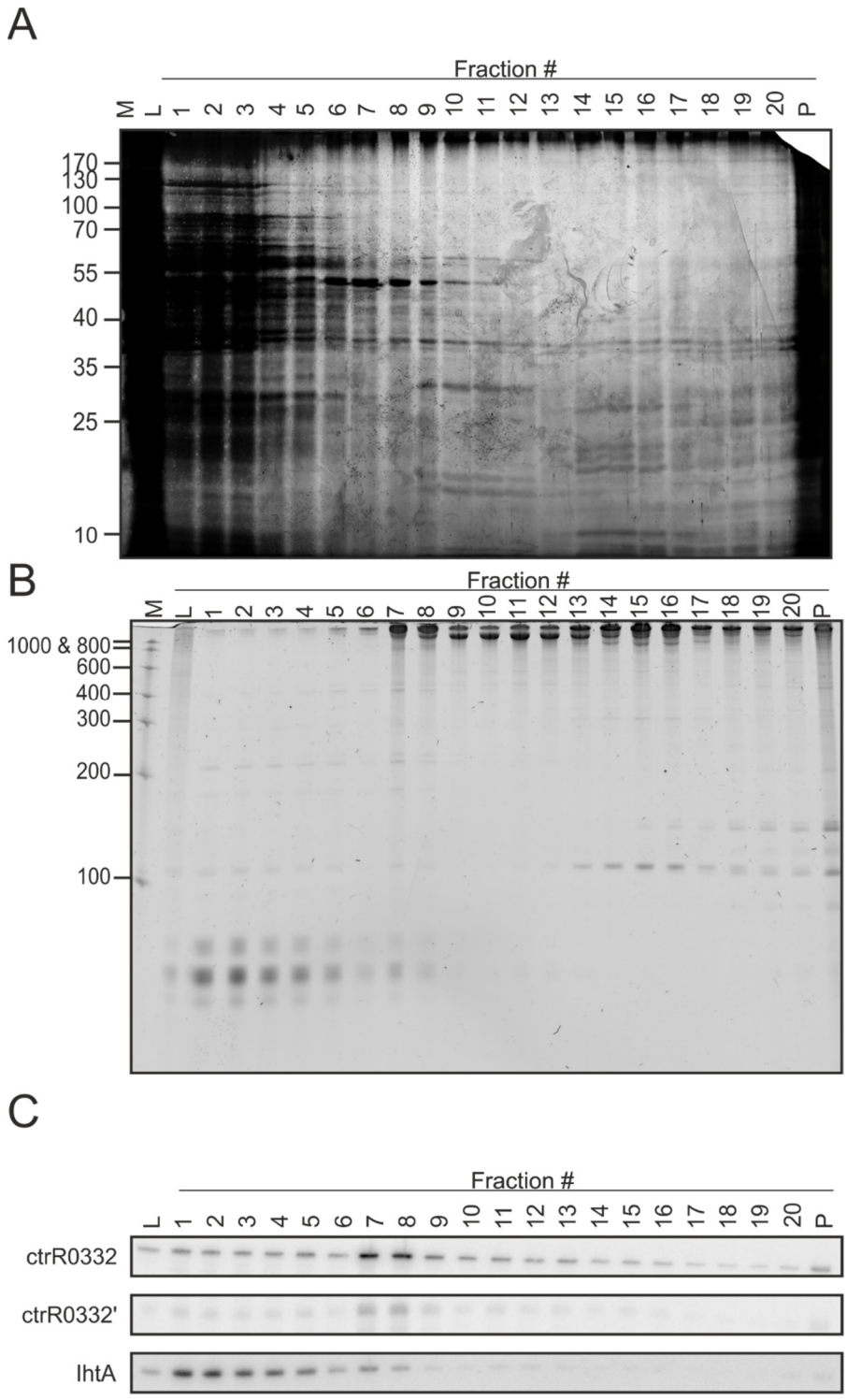


Figure 15: Pellets from thirty 150 cm^2-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep at 5.0 m/s for 20 s once. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. (A) Proteins were separated using a 12% SDS-PAGE and visualised via silver staining. The first lane shows the protein marker (M) followed by a lysis control and the samples from the different fractions of the gradient collected from top (1) to pellet (P). (B) RNAs were loaded onto 7M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows Riboruler Low range, followed by the lysis control and the samples of the different fractions of the gradient collected from top (1) to pellet (P). (C) RNAs were again separated on a 7M Urea 6% PAGE but then transferred to a nylon membrane before probing for the RNAs: ctrR0332 and IhtA. The processed form of ctrR0332 was detected at ~80 bp (ctrR0332'). The order of the samples is as before, but without marker.

To further improve the gradients the speed of the FastPrep was decreased to 4.5 m/s and further to 4 m/s. In addition to quality control via gel, a nanodrop spectrophotometer was used to measure the absorbance at 260 nm for each fraction after centrifugation (Figure 16). At these absorptions a bulk peak, a 30S peak, 50S peak and peak for the pellet in the respective fractions were expected. Although the protein concentrations in the gradients were suitable, their absorption spectra and results of RNA gels indicated an insufficient separation and RNA degradation.

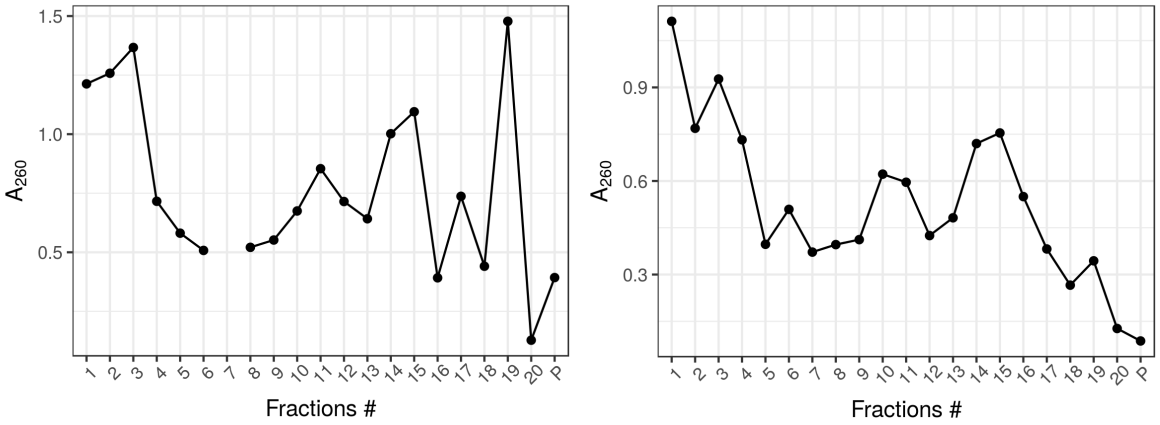


Figure 16: Absorbance of all collected fractions (1-P) measured at 260 nm via NanoDrop spectrophotometer. Pellets of thirty 150 cm^2 dishes of cells infected with *Chlamydia* were used and lysed at (left) 4 m/s for 20 s once and (right) at 4.5 m/s.

For the following gradient, the speed during lysis with the fastprep was 4 m/s. To increase the yield of usable sample of each fraction material of forty 150 cm^2 dishes of *C. trachomatis* infected Hela229 cells were used. Instead of a silver-stained protein gel, coolidal Coomassie staining was used to detect proteins. With that, the bands of the lighter fractions were preserved and ribosomal proteins were seen (Figure 17A). Due to the increased input, signal strenght in RNA gels increased (Figure 17B). Furthermore, the 16S RNA and the 23S RNA separated more clearly.

All nothern blots showed an intense signal in fraction 4, which was due to the high yield resulting from RNA extraction for this fraction. In contrast, fraction 18 seemed to have a lower yield and fraction 11 was not even detectable via northern blot. In addition to the probes for ctrR0332 and IhtA, futher probes for the signal recognition particle (SRP), transfer messenger RNA (tmRNA) and the 5S RNA were included (Figure 17C). CtrR0332 was detectable throughout the gradient, while ctrR0332' intensity was not as present as in the previous

gradient. IhtA and SRP showed a similar course in the gradient with their intensities rising until fraction 4 and then dropping to a faint signal. The signal for tmRNA was abundant in the lysate, and started appearing in fraction 4 (due to the high RNA yield resulting from extraction) and then slowly decreased until no signal was visible in fraction 16. The highest signal intensity of the 5S RNA was detectable in the fractions 12 to 14 while another intensity peak was seen in the light fractions from 1 to 4. 5S RNA was expected to be concentrated in the heaviest fractions of the gradient.

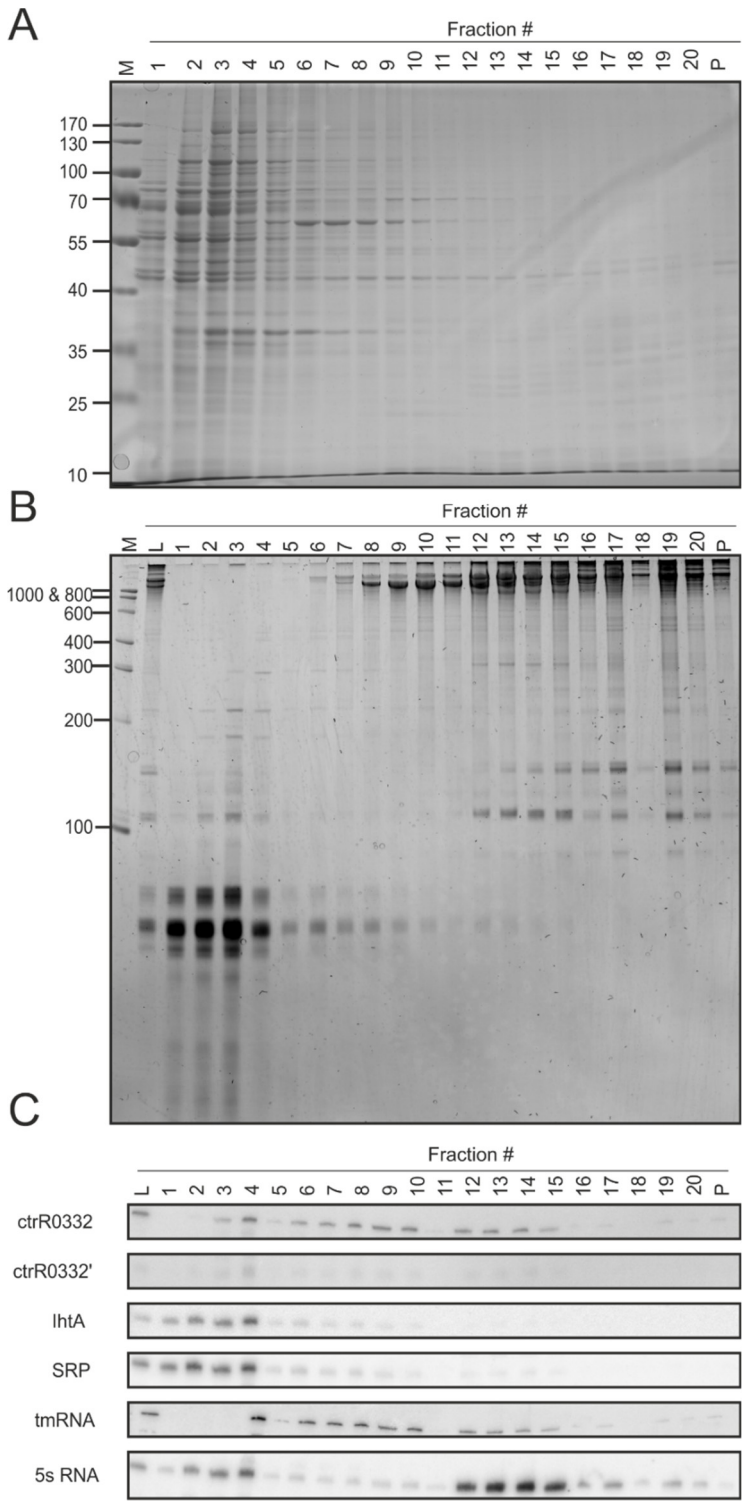


Figure 17: Chlamydial pellets from forty 150 cm^2-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep at 4.0 m/s for 20 s once. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. (A) Proteins were separated using a 12% SDS-PAGE and visualised via colloidal Coomassie staining. The first lane shows the protein marker (M) followed by a lysis control and the samples from the different fractions of the gradient collected from top (1) to pellet (P). (B) RNAs were loaded onto 7M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows RiboRuler Low range, followed by the lysis control and the different fractions of the gradient collected from top (1) to pellet (P). (C) RNAs were again separated on a 7 M Urea 6% PAGE but then transferred to a nylon membrane before probing for the RNAs: ctrR0332 and IhtA, signal recognition particle (SRP), transfer-messenger RNA (tmRNA) and the 5S RNA. The processed form of ctrR0332 was detected at ~80 bp (ctrR0332'). The order of the samples is as described in B, but without marker.

In addition, a nanodrop spectrophotometer was used to measure the absorption (A_{260}) of each fraction. Using this, the general distrubution of the complexes in the gradient can be observed as described in a previous study for *Salmonella* (Smirnov et al., 2016b). The highest peak was seen in fractions 2 and 3. 30S and 50S bulk peaks were mergin into a single peak without showing a strong separation, while the pellet or 70S peak were not as strong as described in the orignal method and were only a fraction of the 50S (Figure 18).

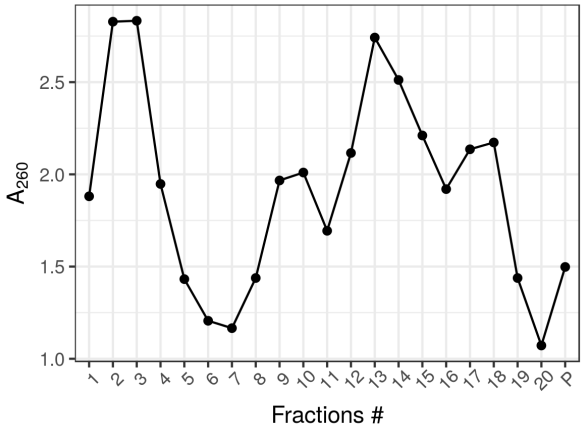


Figure 18: Absorbance of all collected fractions (1-P) measured at 260nm via NanoDrop spectrophotometer. Pellets of forty 150 cm² dishes of cells infected with *Chlamydia* were used and lysed at 4 m/s for 20 s once. The lysate was separated using a 10%-40% glycerol gradient before samples were taken.

To improve the separation, the latest described conditions were repeated twice with similar results (Appendix figure 39 to Appendix figure 42). The absorption spectra indicated a similar distribution of the complexes in these gradients.

For the last gradient performed with the FastPrep, the lysis time was decreased to 15 s at 4 m/s. The decreased processing time resulted in a gradient, which was closest to the ideal separation as described in the original Grad-seq publication (Smirnov et al., 2016b) according to A_{260}-values of the fractions after fractionation. There was a clear bulk peak in fraction 2, a peak representing the 30S subunit in the fractions 10 to 12, a 50S subunit peak from fraction 14 to 17 and a major increase in absorption in the pellet (Figure 19).

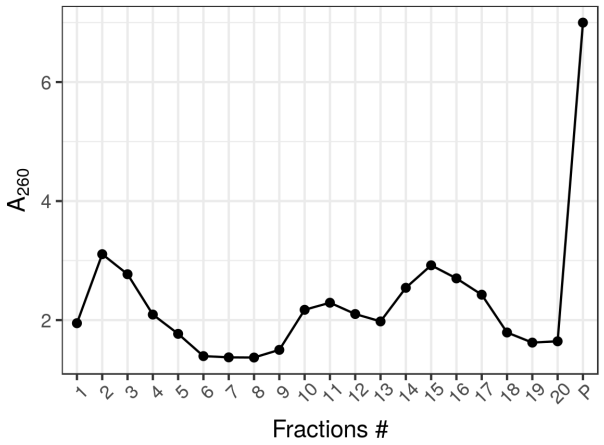


Figure 19: Absorbance of all collected fractions (1-P) measured at 260nm via NanoDrop spectrophotometer. Pellets of forty 150 cm² dishes of cells infected with *Chlamydia* were used and lysed with a fastprep at 4 m/s for 15 s once. The lysate was separated using a 10%-40% glycerol gradient before samples were taken.

The colloidal Coomassie stained protein gel and the RNA gel reflected the results from the absorption measurement. The RNA gel showed the tRNAs in the light fraction, where the bulk peak resides and no big complexes are expected. The fractions of the RNA gel for 16S RNA and 23S RNA coincide with the peaks from the absorption measurement (Figure 20B). Additionally, this gradient included a lysis control for the protein samples to estimate if enough protein was present for separation and to observe the general distribution of proteins after lysis. GroEL was distinctly observed at ~60 kDA from fraction 7 to fraction 12. The ribosomal subunit alpha was detected at ~42 kDa in the fractions 1 to 4. Ribosomal proteins were expected to be in the heavy fractions in the molecular light regions of the gel (Figure 20A). In addition to the previously used northern blot probes, which had shown a similar distribution across the gradient, we included two probes for the human host, since cultivation of *Chlamydia* was done in Hela229 cells. The previously used probes showed the same characteristics as in the last gradient. The human 5S RNA was concentrated in the heaviest fractions of the gradient and the 5.8S RNA shared this distribution. The chlamydial and the human 5S RNA shared both signals in the light fractions of the gradient (Figure 20C).

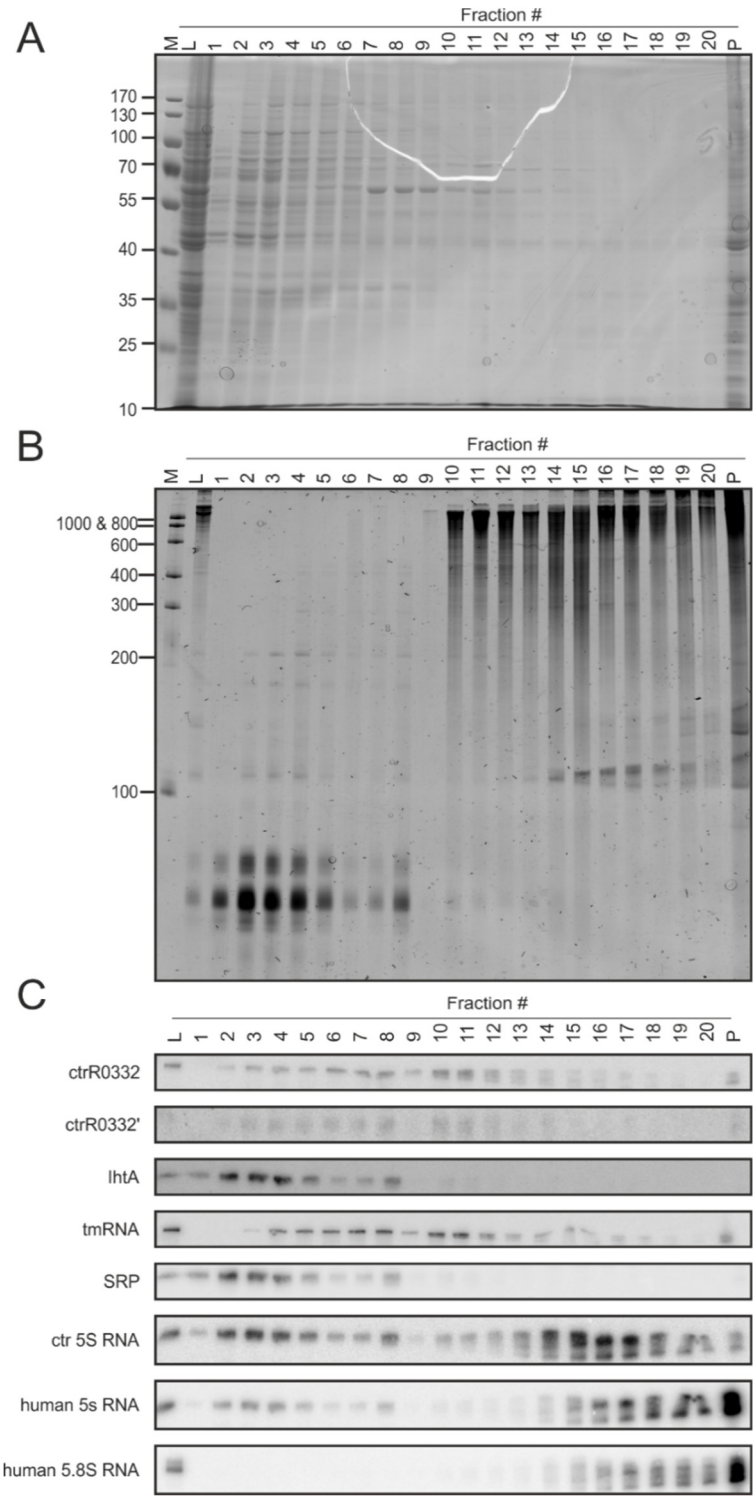


Figure 20: Chlamydial pellets from forty 150 cm^2-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep at 4.0 m/s for 15 s once. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. (A) Proteins were separated using a 12% SDS-PAGE and visualised via colloidal coomassie staining. The first lane shows the protein marker (M) followed by a lysis control and the samples from the different fractions of the gradient collected from top (1) to pellet (P). (B) RNAs were loaded onto 7M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows RiboRuler Low range, followed by the lysis control and samples from the different fractions of the gradient collected from top (1) to pellet (P). (C) RNAs were again separated on a 7 M Urea 6% PAGE but then transferred to a nylon membrane before probing for the RNAs: ctrR0332 and IhtA, signal recognition particle (SRP), transfer-messenger RNA (tmRNA), chlamydial 5S RNA, human 5S RNA and human 5.8S RNA. The processed form of ctrR0332 was detected at ~80 bp (ctrR0332'). The order of the samples is as described in B, but without marker.

Instead of using the FastPrep for lysis preparation, in a next step, 0.1 mm silica beads were used. This procedure included 15 s of vortexing and 15 s of cooling for 5 times before loading the samples onto the glycerol gradient prepared with a gradient master station. This resulted in similar absorption spectra as the gradient performed at 4 m/s for 15 s with the FastPrep. The bulk peak was clearly identifiable at fraction 2 and 3, decreasing slowly to the lowest value at fraction 8. This was followed by the 30S RNA peak from the fractions 10 to 13, and the stronger 50S subunit peak from fraction 14 to 17. The gradient ended with high absorption in the pellet fraction (Figure 21).

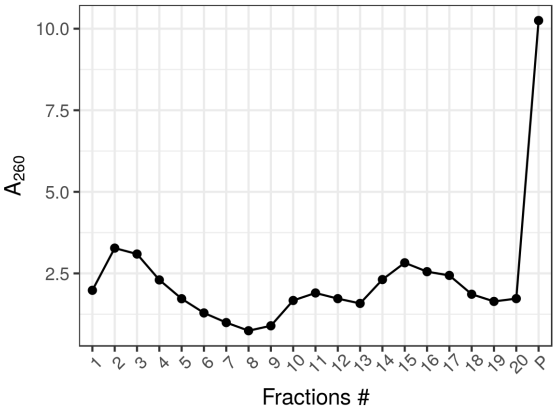


Figure 21: Absorbance of all collected fractions (1-P) measured at 260nm via NanoDrop spectrophotometer. Pellets of forty 150 cm^2 dishes of cells infected with *Chlamydia* were lysed using 0.1 mm silica beads by 15 s of vortexing and 15 s of cooling on ice for 5 times. The lysate was separated using a 10%-40% glycerol gradient before samples were taken.

The protein gel clearly showed a band representing GroEL at ~ 60 kDa in the middle of the gradient, where medium sized complexes are expected and a band representing RpoA at ~40 kDa in the light fractions of the gradient where small complexes are expected. Ribosomal proteins are usually small but due to forming complexes they were observed in the heavy fraction (Figure 22A). The RNA gels showed the same results as the previously performed RNA gels of the other gradients (Figure 22B). In addition, the results of the northern blots were similar to the results of northern blots from the gradients as shown above. This indicates that the two used methods give the same results (Figure 22C). This method of lysis was replicated twice with identical results (Appendix figure 43 and Appendix figure 44).

After incremental steps of fine tuning the lysis of chlamydial and adjustments of sample input, the last-mentioned conditions were identified as ideal for sample preparations for high-throughput methods (e.g. RNA-seq and MS of each fraction).

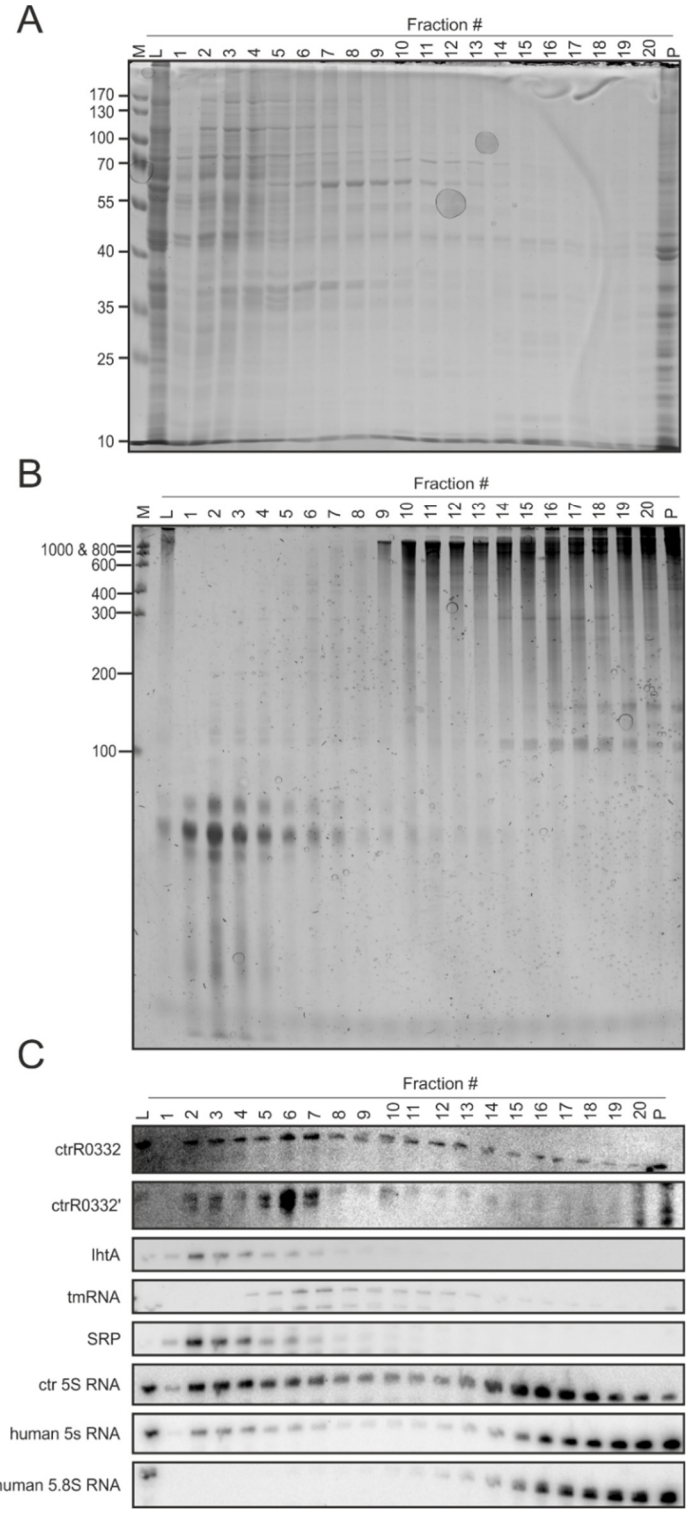


Figure 22: Chlamydial pellets from forty 150 cm^2-dishes of infected Hela229 cells were lysed using 0.1 mm silica beads by 15 s of vortexing and 15 s cooling on ice for 5 times. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. (A) Proteins were separated using a 12% SDS-PAGE and visualised via colloidal Coomassie staining. The first lane shows the protein marker (M) followed by a lysis control and the samples from the different fractions of the gradient collected from top (1) to pellet (P). (B) RNAs were loaded onto 7 M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows Riboruler Low range, followed by the lysis control (L) and samples from the different fractions of the gradient collected from top (1) to pellet (P). (C) RNAs were again separated on a 7 M Urea 6% PAGE but then transferred to a nylon membrane before probing for the RNAs: ctrR0332 and IhtA, signal recognition particle (SRP), transfer-messenger RNA (tmRNA), chlamydial 5S RNA, human 5S RNA and human 5.8S RNA. The processed form of ctrR0332 was detected at ~80 bp (ctrR0332'). The order of the samples is as described in B, but without marker.

3.3. Analysis of Grad-seq

In this study, high-throughput methods were used to globally screen for complexes formed in *Chlamydia*. The remaining samples of the gradient, with the lysis performed by vortexing instead of preparation with the FastPrep, were used as input. The RNA samples were supplemented with ERCC spike-ins before sequencing. The protein samples were supplemented with UPS2 proteomic standard before homogenises and MS (see Preparation of RNA and protein samples for high-throughput analysis, p. 40).

The output per library was adjusted according to the RNA species that were expected and to the absorption spectrum (Figure 21), so that the rRNAs do not overshadow other RNA species. More reads than requested were obtained and FastQC revealed that most of the reads were of good quality. After trimming with cutadapt (Grad-seq trimming, p. 135) less than 1% of the reads were removed, confirming the high-quality output during sequencing (Table 21).

Table 21: Distribution of reads per library before and after quality control (QC).

Fraction	requested in Millions	reads pre-QC	reads post-QC	reads survived after QC (%)
00L	20	28217560	28145158	99.74
1	15	29129946	29063012	99.77
2	10	14346244	14319563	99.81
3	10	12443994	12421285	99.82
4	7	9342322	9322315	99.79
5	7	9474507	9461128	99.86
6	7	8036626	8022406	99.82
7	7	7767417	7747658	99.75
8	7	7493274	7483196	99.87
9	10	11243881	11224487	99.83
10	25	26459531	26394149	99.75
11	30	31384063	31310467	99.77
12	25	25767821	25723322	99.83
13	20	21466558	21426256	99.81
14	35	37221256	37159546	99.83
15	40	40080348	39964045	99.71
16	35	37836544	37773935	99.83
17	35	35951755	35886506	99.82
18	25	27024276	26973934	99.81
19	20	21302461	21264831	99.82
20	20	21489660	21452942	99.83
21P	15	17387381	17362497	99.86

The trimmed reads were then used in the READemption pipeline (Förstner et al., 2014) to perform mapping using the segemehl mapper (Hoffmann et al., 2009) and gene quantification with the build-in subcommand of the pipeline. On average, 92% of our input reads were mapped to the human reference genome (GRCh38.p10), the chlamydial reference genome (Genome: AM884176.1, Plasmid: AM886278.1) and the sequences of the ERCC spike-in. Towards the heavy fractions, the amount of uniquely aligned reads dropped significantly. As expected, these fractions are mainly composed of ribosomal transcripts, since ribosomal depletion was not performed. The light fractions proportionally showed more uniquely mapped reads indicating a wider variety of transcripts (Table 22).

Table 22: Mapping statistics from the RNA-seq for the 22 libraries of the gradient.

Libraries	No. of input reads	No. of reads - Removed as too short	No. of reads - Long enough and used for alignment	Total no. of aligned reads	Total no. of unaligned reads	Total no. of uniquely aligned reads	Total no. of alignments	Percentage of aligned reads (compared to no. of input reads)	Percentage of aligned reads (compared to no. of long enough reads)	Percentage of uniquely aligned reads (in relation to all aligned reads)
00L	28145158	32317	28112841	26745853	1366988	4998595	91140095	95.03	95.14	18.69
1	29063012	212419	28850593	25438651	3411942	21585449	36469854	87.53	88.17	84.85
2	14319563	45695	14273868	13032048	1241820	7882594	22904884	91.01	91.3	60.49
3	12421285	33636	12387649	11428036	959613	6204216	22112843	92	92.25	54.29
4	9322315	36775	9285540	8553330	732210	4751587	17042703	91.75	92.11	55.55
5	9461128	12474	9448654	8825265	623389	3815581	17947748	93.28	93.4	43.23
6	8022406	16839	8005567	7549256	456311	3000866	15449222	94.1	94.3	39.75
7	7747658	23312	7724346	7245327	479019	2790188	15017323	93.52	93.8	38.51
8	7483196	13338	7469858	7091778	378080	2405586	15166607	94.77	94.94	33.92
9	11224487	12041	11212446	10849876	362570	2766380	23212287	96.66	96.77	25.5
10	26394149	15084	26379065	25656331	722734	3565543	54396975	97.2	97.26	13.9
11	31310467	20656	31289811	30413978	875833	4504340	67496507	97.14	97.2	14.81
12	25723322	10489	25712833	25106381	606452	4104953	60992228	97.6	97.64	16.35
13	21426256	9599	21416657	20889682	526975	2193262	51880401	97.5	97.54	10.5
14	37159546	14684	37144862	36280146	864716	2206424	85799420	97.63	97.67	6.08
15	39964045	27430	39936615	38830433	1106182	1412293	98490436	97.16	97.23	3.64
16	37773935	15133	37758802	36723880	1034922	1120301	112844813	97.22	97.26	3.05
17	35886506	15697	35870809	34840597	1030212	1158632	116110278	97.09	97.13	3.33
18	26973934	12730	26961204	25965899	995305	900526	96830119	96.26	96.31	3.47
19	21264831	8453	21256378	20309655	946723	864770	102887320	95.51	95.55	4.26
20	21452942	11028	21441914	20462498	979416	1113878	134115448	95.38	95.43	5.44
21P	17362497	9375	17353122	16480878	872244	1534874	116699203	94.92	94.97	9.31

The distribution of the coverage across the genome for the single fractions showed that the lighter fractions and the lysis control had a higher percentage of regions covered in the genome with a sequencing depth up to 400. In contrast, heavier fractions from the bottom of the gradient tend to have percentwise only a few regions with a coverage up to 400 (Figure 23). For these fractions few regions in the genome with a very high coverage were found.

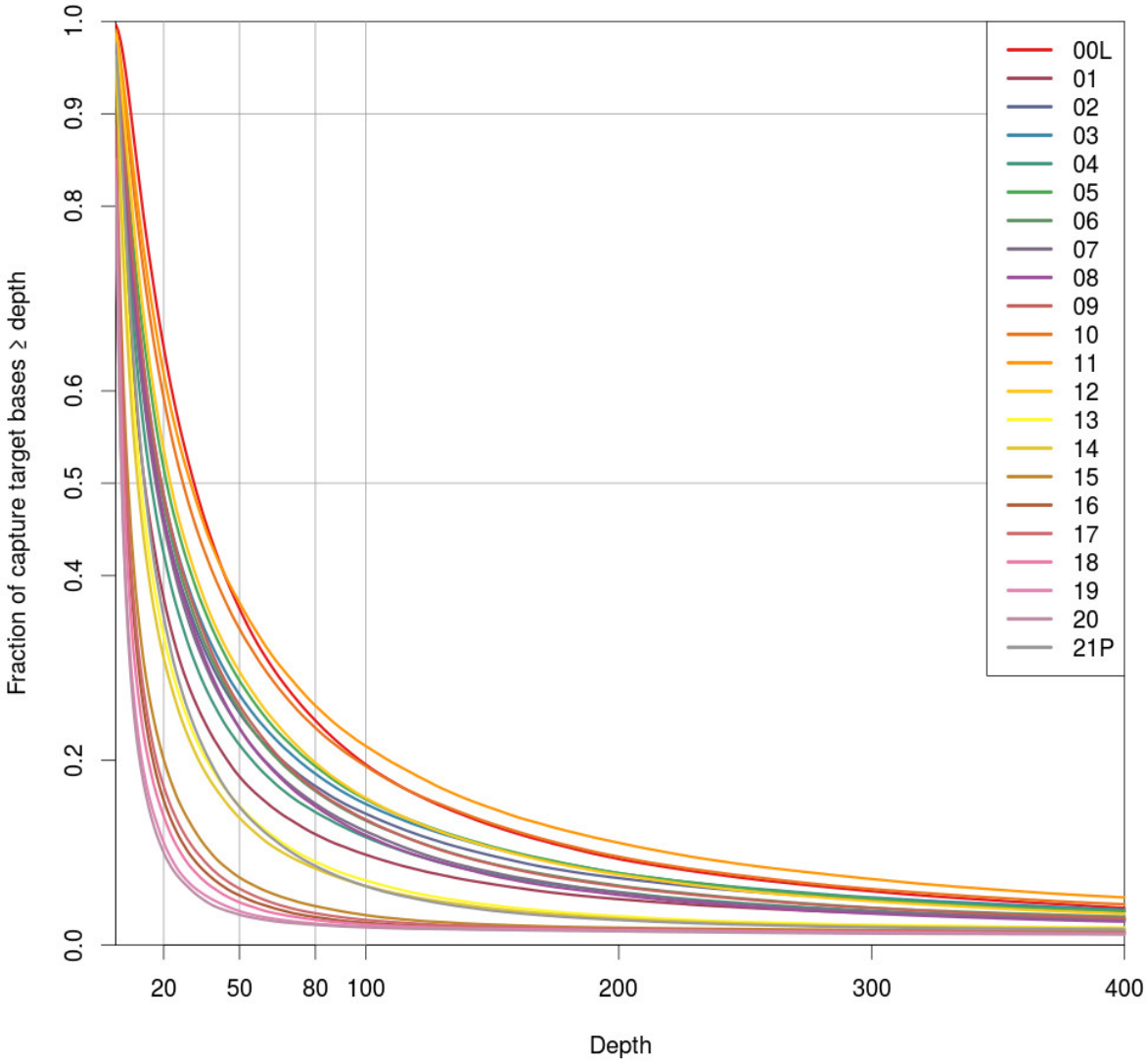


Figure 23: Coverages of the references per fraction. The read depth is plotted against the fraction of targeted bases (e.g. the genome references and the ERCC spike-in sequences).

The gene quantification build-in from READemption was used to count the reads per annotation. Therefore, the reference annotation of the genomes and the annotation from the TagRNA-seq analysis were used. The number of quantified reads corresponded to the number of reads mapped in the libraries. The lysis control represents the distribution of RNA classes in *Chlamydia,* with rRNAs being the most abundant transcripts. In the gradient, the most abundant RNA class was rRNAs, except for the first fraction which was dominated by tRNAs.

The number of tRNA reads decreased with increasing fraction number, while the number of reads of rRNAs increased from fraction two on, dominating the other RNA species. The other RNA classes all slowly decreased with increasing fraction number. Human transcripts (hCDS) started to increase in the late fractions again. Chlamydial coding transcripts (CDS) peaked in fraction 11 (Figure 24).

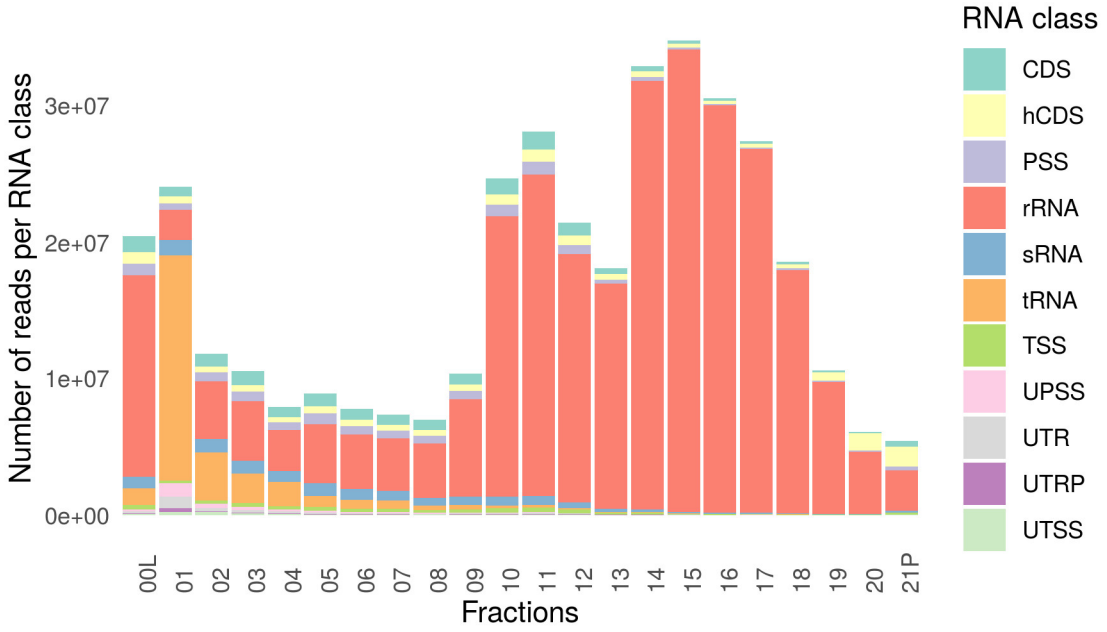


Figure 24: Number of reads per RNA class per sequencing library from the reference annotation with addition of the TagRNA-seq established new annotation.

The different sequencing depths of the libraries were corrected with the statistics from the ERCC spike-ins as references. Quantified transcripts were then normalised to their maximum value and represented in heatmaps.

In a global representation of the chlamydial RNAs and proteins, it was seen that based on the observed type of molecule different distributions were measured. RNAs tended to be more focused in the top half of the gradient while proteins were more spread throughout the gradient. The RNAs from *Chlamydia* were abundant in fractions 2 to 12, with numerous RNAs peaking in fraction 6 and 11. Furthermore, many RNAs were present in the pellet of the gradient (Figure 25A). On the other hand, proteins were distributed across the gradient with the fractions 17 to 20 only having a few quantified proteins (Figure 25B). Most of the proteins were present in the light fractions 1 to 4, as well as in the pellet fraction. In contrast, many RNAs also appeared in the middle fractions of the gradient.

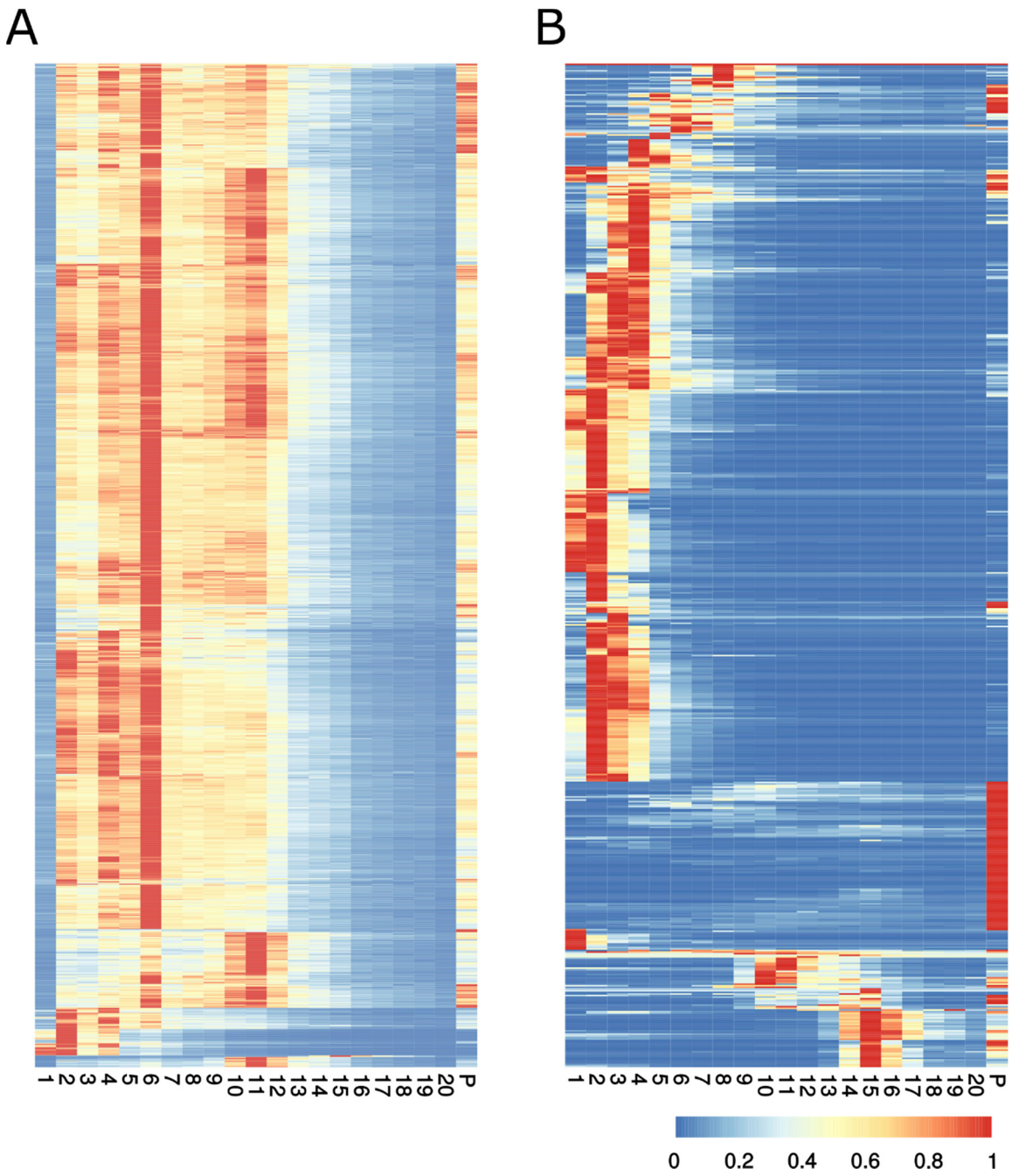


Figure 25: Quantified RNAs/Proteins represented as heatmaps normalised to their maximum value. Red indicates high abundance with a value close to 1, while blue represents a value close to 0 indicating low abundance. (A) Representation of RNAs from the reference annotation and the annotation from TagRNA-seq analysis. (B) Proteins from mass spectrometry analysis for all detectable chlamydial proteins.

Using a principle component analysis (PCA), a clear separation between chlamydial proteins (green in Figure 26A) and the coding RNAs (red in Figure 26A) was seen. The proteins were themselves split into two clusters. This separation was not dependent on protein function or structure. While tRNAs were exclusively present in one of the protein's clusters, rRNAs were present in both protein clusters. sRNAs were present in the clusters formed by coding RNAs

and proteins, with the most of them being present in the border area between these two clusters (Figure 26A).

Using only the RNAs from the reference annotation and the generated TagRNA-seq annotation, most of the quantified RNAs were seen in one cluster, while the tRNAs formed their own cluster. The known sRNAs were present in the fringes of the cluster that was formed by the majority of the RNAs.

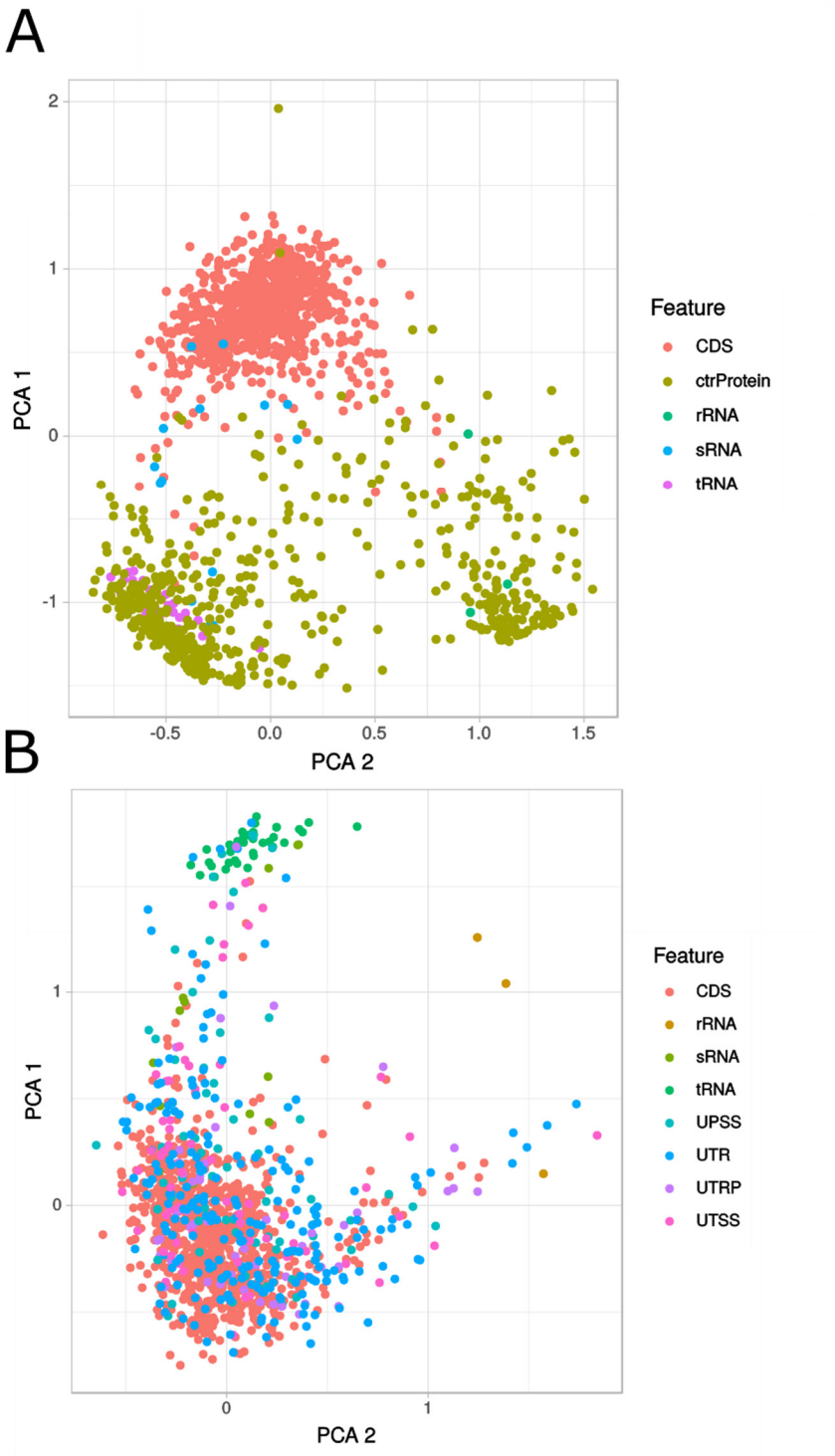


Figure 26: Principle component analysis (PCA) of the first two principle components of the quantified and normalized molecules from the Grad-seq approach. (A) PCA of the reference annotation of the RNAs and the proteins. (B) PCA of reference RNAs and the RNA annotation from the TagRNA-seq analysis.

Heatmaps were used to visualize the presence of different rRNAs (Figure 27A) in the different fractions. A similar distribution of rRNAs as in the northern blots and RNA gel analyses was observed (Figure 22B and C). In the reference annotation, each RNA was doubly present in the genome. The quantification between the two rRNA copies showed no difference. tRNAs were

mostly present in the light fractions of the gradient, which was also confirmed by the RNA gel of the gradient (Figure 22B). Although tRNAs have multiple copies the copies themselves have different characteristics. Most of the tRNAs were highly abundant in the second fraction and many of them also had a second peak in fraction 4. Furthermore, a few tRNAs were seen in fraction 6. Neither rRNAs nor the tRNAs were present in the pellet (Figure 27).

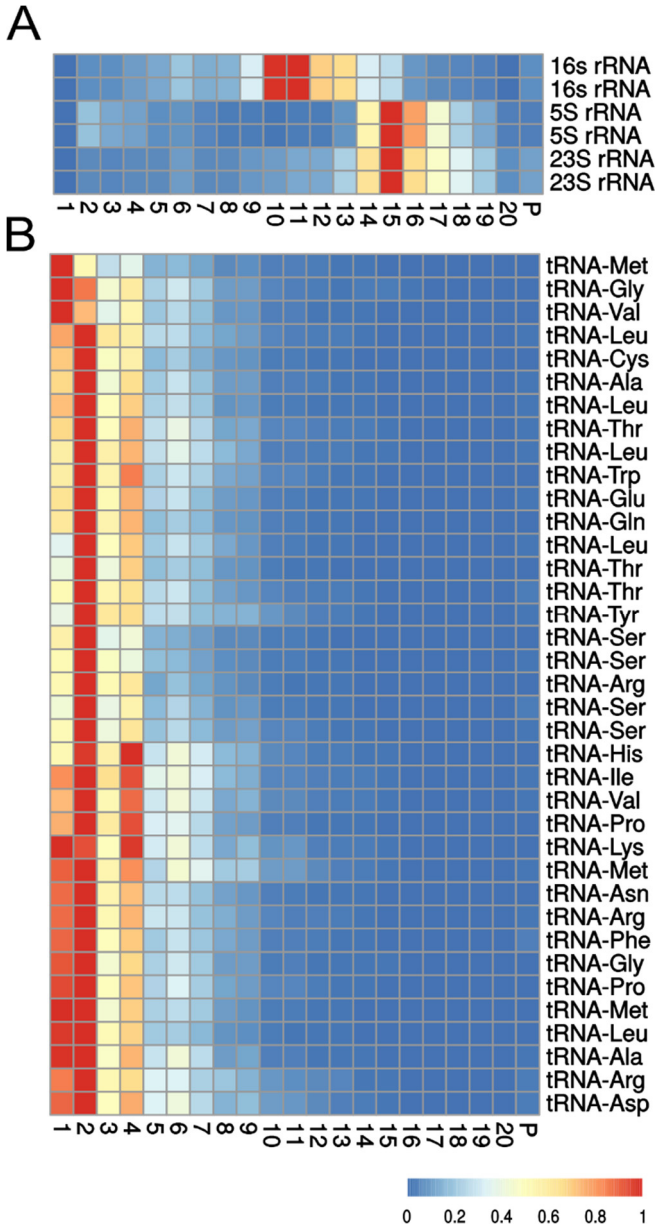


Figure 27: Quantified RNAs from the reference annotation represented as heatmaps normalised to their maximum value. Red indicates high abundance with a value close to 1, while blue represents a value close to 0 indication low abundance. (A) Quantification of ribosomale RNAs. (B) Quanification of tRNAs.

Heat map analysis of the intergenic annotation from TagRNA-seq revealed that these regions were also quantifiable and that the majority of the RNAs had highest values in fraction 6 (Figure 28). The intergenic transcripts had peaks in the fractions 2 and 6 (Figure 28A and C) while the UTR were highly abundant in fractions 6 and 11 (Figure 28B and C).

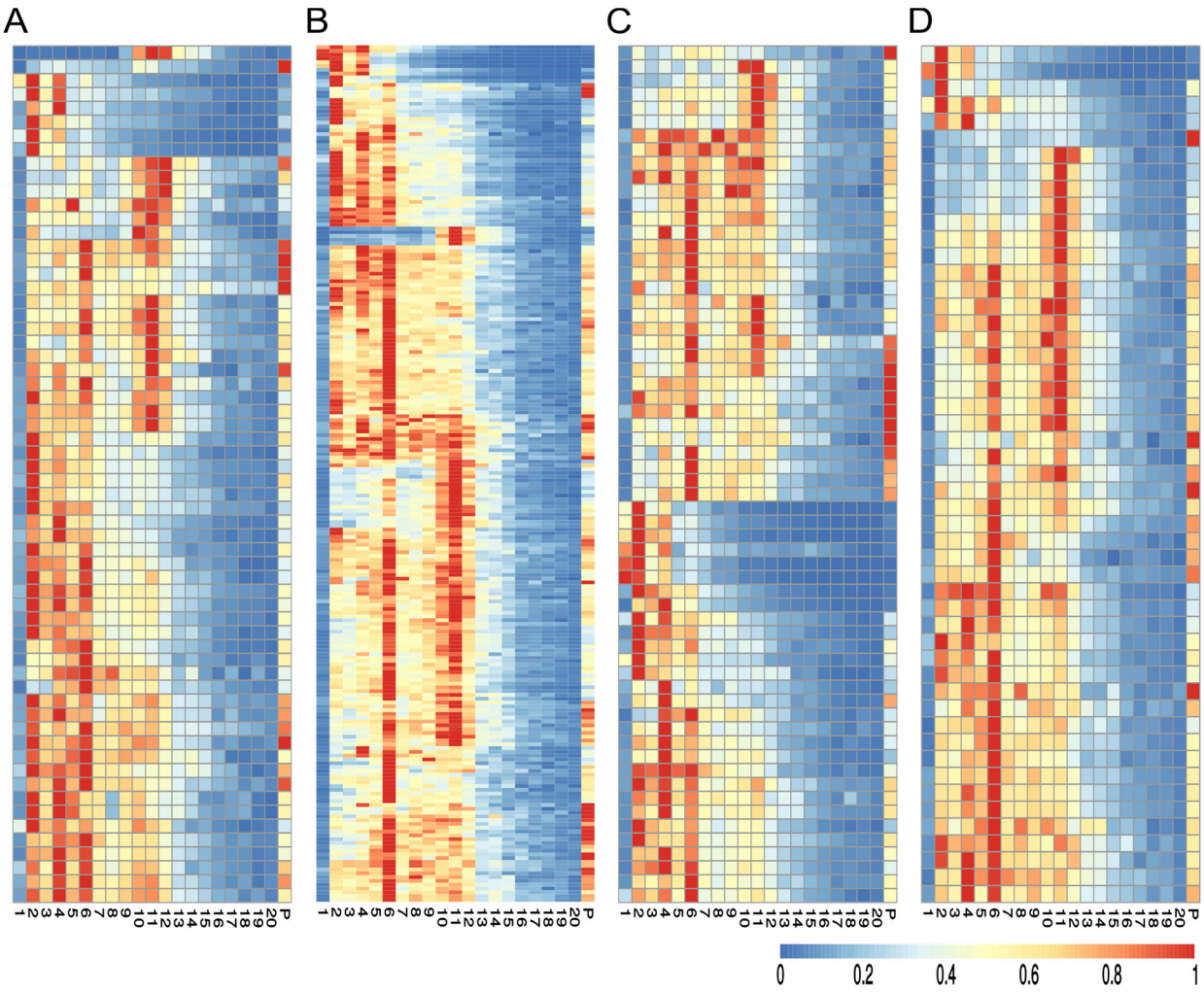


Figure 28: Quantified intergenic annotations from the TagRNA-seq analysis represented as heatmaps normalized to their maximum value. Red indicates high abundance with a value close to 1, while blue represents a value close to 0 indication low abundance. (A) Quanitfication of the unknown intergenic transcriptional start sites. (B) Heatmap of the the quanified 5'-untranslated regions of RNAs. (C) Quanitfication of the unknown intergenic processing sites. (D) Heatmap of the the quanified 5'-untranslated regions of RNAs orginating from processing events.

Analysis of rRNAs encoding ribosomal proteins as well as quantification of the ribosomal proteins themselves showed a vastly different picture. The rRNAs were abundant in the first half of the gradient with high values in fraction 6 and 11 (Figure 29A). In contrast, ribosomal proteins were exclusively present in the fractions with rRNAs indicating that the ribosomal complexes are intact (Figure 29B,Figure 27A). Four proteins which associate with the ribosome were only present in the very light fractions of the gradient. These are: 50S ribosomal protein L7/L12 (*rpl*L), Ribosome-binding factor A (*rbf*A), Ribosome-binding ATPase (*ych*F) and Elongation factor G (*fus*A).

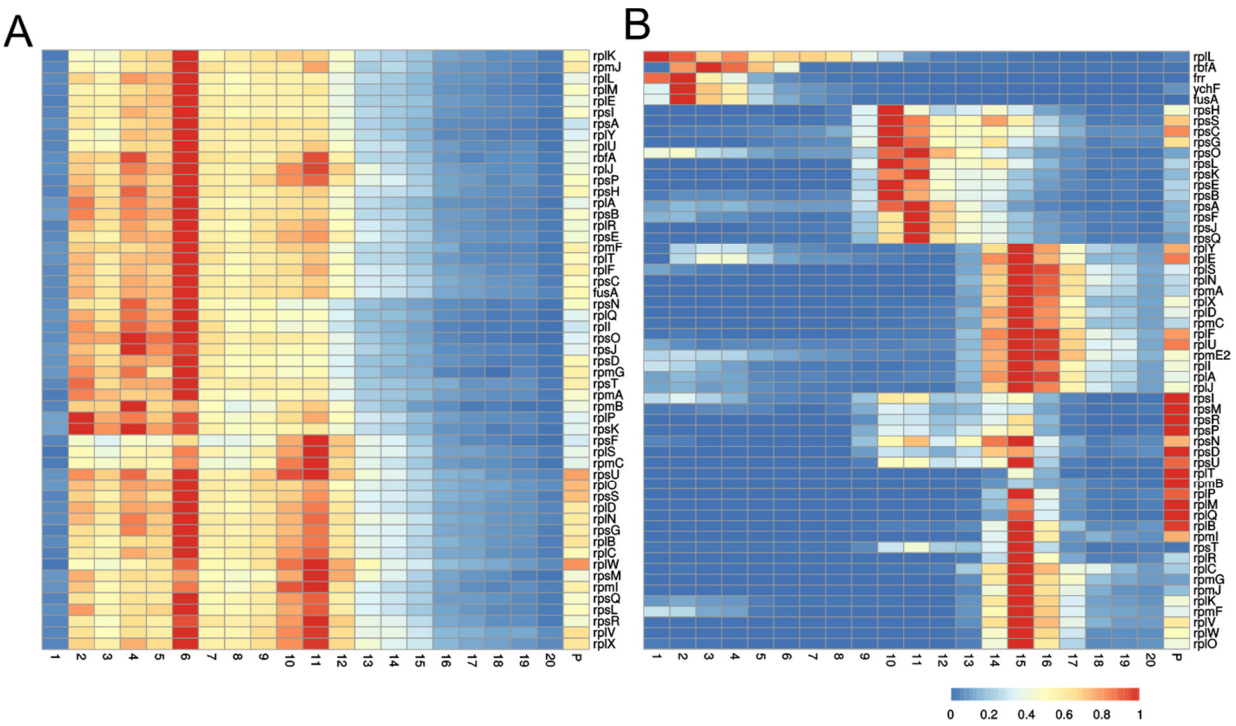


Figure 29: Quantifiction of the RNAs encoding ribosomal proteins from (A) the RNA-seq of the Grad-seq approach and (B) quanification of the ribosomal proteins by mass spectrometry represented as heat maps normalized to their maximum value. Red indicates high abundance with a value close to 1, while blue represents a value close to 0 indication low abundance.

The RNAs encoding σ-factors and the RNAP subunits were abundant in fraction 6 (Figure 30A), while the associated proteins show a different fractional distribution (Figure 30B). The σ-factors are present in the first two fractions, while the RNAP subunits are mostly in fraction 6 and the pellet. The alpha subunit is spread across several fraction (2-8) (Figure 30).

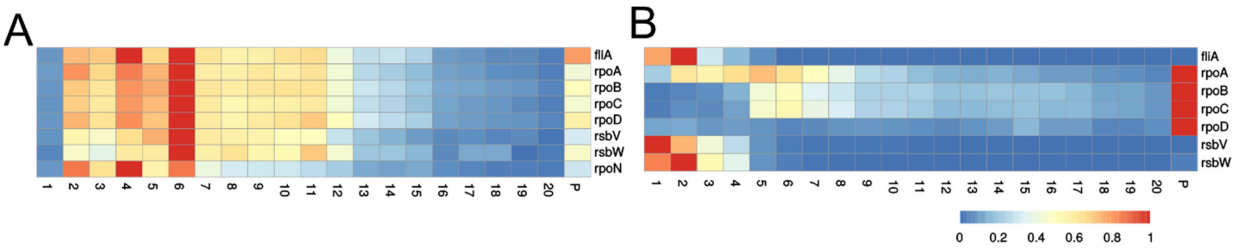


Figure 30: Quantifiction of the RNAs encoding σ-factors and RNA polymerase from (A) the RNA-seq of the Grad-seq approach and (B) quanification of the proteins encoded by the RNAs by mass spectrometry represented as heat maps and normalized to their maximum value. Red indicates high abundance with a value close to 1, while blue represents a value close to 0 indication low abundance.

Analysis of ncRNAs revealed that these RNAs had their highest values in the same fractions as previously observed for other RNAs (2, 4 and 6). The sRNA IhtA acts by direct base pairing and is believed to not require a chaperon. IhtA was only present in the fractions 2 to 4. The sRNA identified by Albrecht and colleagues, ctrR0332 (Albrecht *et al.*, 2011, Albrecht *et al.*, 2010), was highly present in fractions 2 to 12. This RNA has a processing site generating a ~80 bp

(ctrR0032f) and a ~160 bp (ctrR0032b) transcript. Overall, ctrR0032b was not as abundant as ctrR0032f being present also in the fractions 8 to 11 (Figure 31A). In the pellet fraction, only the sRNA ctrR8 was detected. tmRNA and SRP appeared in the same fraction as in the northern blots. The protein of Ctl0077 peaked in fraction 4 with the respective RNA peaking in fraction 6. These are the same fractions in which ctrR0332 shows its maximum abundance. The RNA of ctl0077 and sRNA ctrR0332 have a similar slope in the gradient, while the proteins decrease very rapidly and are almost undetectable in fraction 12 (Figure 31B).

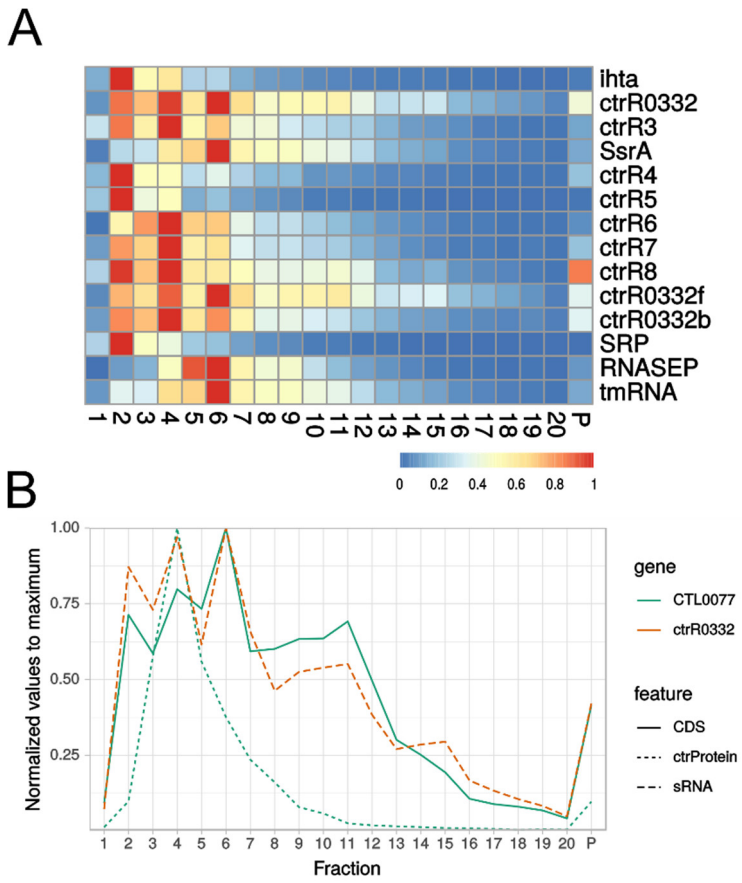


Figure 31: (A) Quantification of known noncoding RNAs from Grad-seq represented as heat map. Red indicates high abundance with a value close to 1, while blue represents a value close to 0 indication low abundance. (B) Vizualization of the possible complex of the small RNA ctrR0332 and the protein and RNA of ctl0077.

3.4. Validation of complexes by *In vitro* oligo aptamer pull-down

The oligo aptamer pull-down was performed to validate the results of the sRNA ctrR0332 – ctl0077 interaction, which was detected via gradient profiling. Furthermore, the analysis was done to identify a possible interaction between IhtA and proteins. Grieshaber and co-workers postulated in 2006 that IhtA may be a trans-encoded antisense sRNA, having no Sm-like protein as Hfq to mediate its RNA-RNA interaction (Grieshaber et al., 2006a).

The initial pull-down was performed with pellets obtained from 120 petri dishes (150 cm^2) of Hela229 cells infected with *Chlamydia* (20 dishes/RNA bait). The pull-down was performed as described above (Protein capture via oligo aptamer RNAs, p. 42). After lysis, an input control was taken. In addition to the RNAs of interest, the reverse complement of the adapter bound to the beads was used as a negative control. Positive controls were RNAseP, tmRNA and 5S RNA from *C. trachomatis*. IhtA and CtrR0332 were the RNAs of interest. To test whether proteins were successfully captured via the RNA baits, the samples were loaded onto to a 12%-SDS PAGE gel and visualized via silver staining (Figure 32). Compared to the rest of the samples, only a few faint bands were observed for the negative control. In all other samples clearly distinguishable bands were detected. RNAseP had the largest protein band at ~130 kDa. For IhtA, a strong band at 25 kDA was detected and not present in the other samples. In the lane of ctrR0332, a protein band at 130 kDA was observed. The tmRNA had a similar band pattern as the negative control, but with higher signal intensity. IhtA and ctrR0332 lacked the band at ~20 kDa, which was present in the lanes for RNAseP, tmRNA and 5S RNA control.

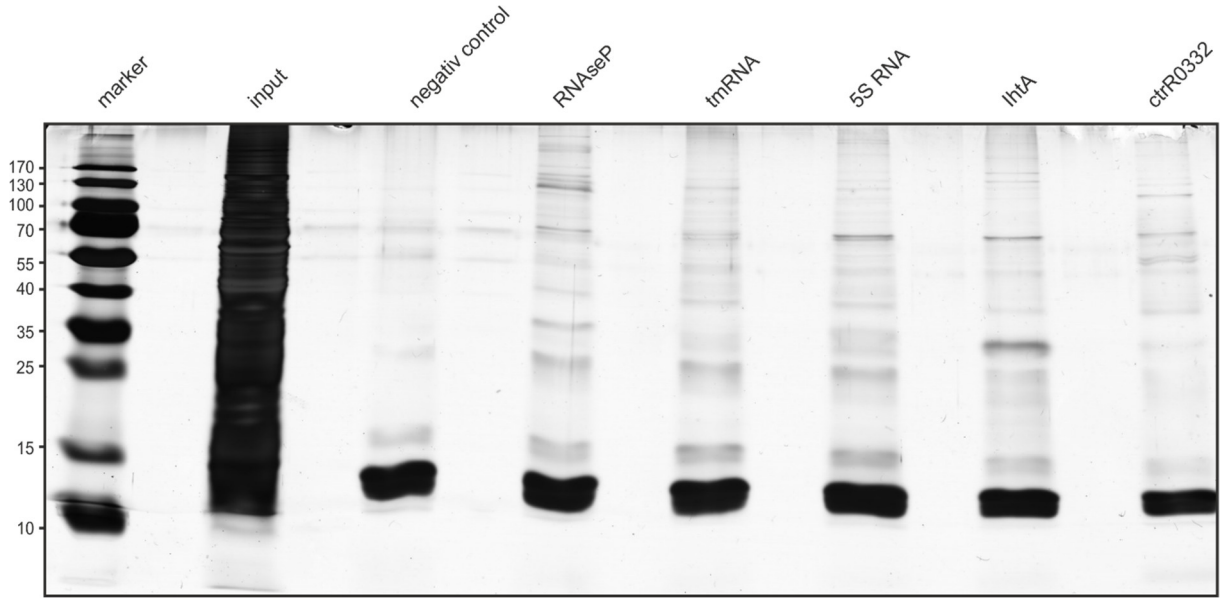


Figure 32: Silver stained 12 % SDS-PAGE-Gel with samples of the oligo aptamer pull-down. Loaded from left to right with one empty lane between each sample: marker, input after lysis, negative control bait, RNAseP bait, tmRNA bait, 5S RNA bait, IhtA bait and ctrR0332 bait. Per bait, twenty 150 cm^2 dishes of Hela229 cells infected with *C. trachomatis* L2/434/bu were used.

A second pull-down was performed from thirty 150cm^2 petri dishes of *Chlamydia* infected cells per RNA bait. This second time, only the negative control, 5S RNA, IhtA and ctrR0332 baits were used. To further increase the material concentration, no gel was made, and the samples were directly prepared for MS. RNaseP and tmRNA were excluded because the baits did not

capture their specific interaction partners. The MS results from both capture assays were statistically analysed. All samples captured pnp (Polyribonucleotide nucleotidyltransferase) and UvrD, a helicase with DNA-dependent ATPase activity. The 5S RNA captured significantly more ribosomal proteins compared to the negative control (Figure 33).

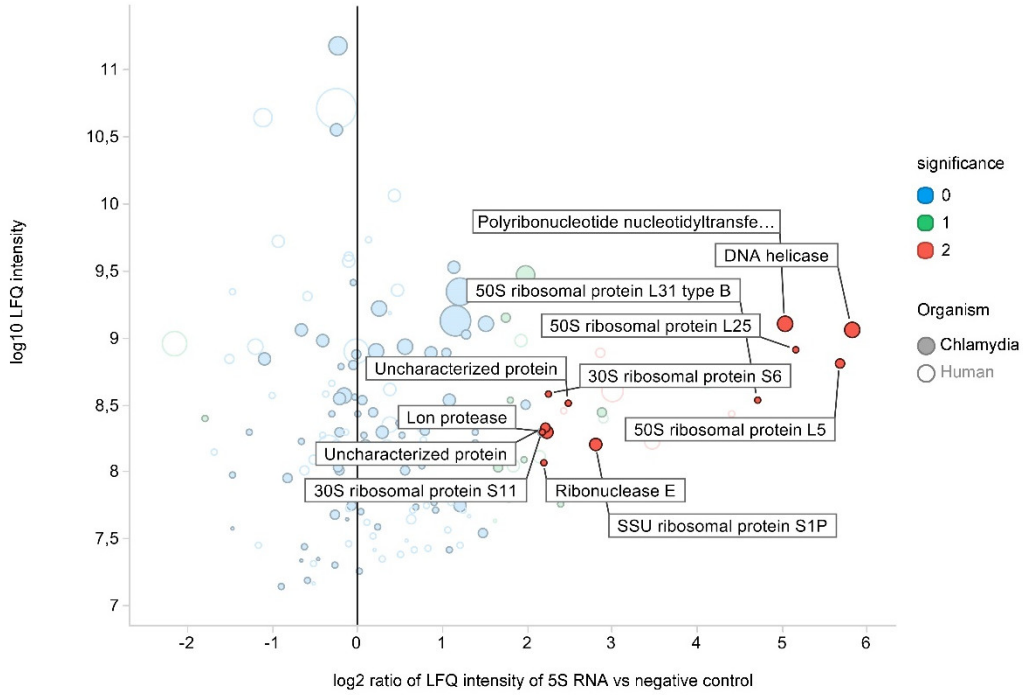


Figure 33: Results of the oligonucleotide pull-down with 5S RNA compared to the results from the negative control. The Y-axis indicates the intensity of the signal and the X-axis shows the ratio between 5S RNA and the negative control. The further right a signal is the more it is enriched compared to the negative control. The size of the signals indicates how many unique razor entries were found. The colour indicates the significance of the signal.

The pull-down with IhtA had no significantly enriched proteins besides pnp and UvrD (Figure 34), indicating that no unique interaction partner were present. In addition to the chlamydial proteins, host proteins were present in the samples. These were ignored, because the RNA baits should not be in contact with the host under natural circumstances. Therefore, these interactions were categorized as unspecific.

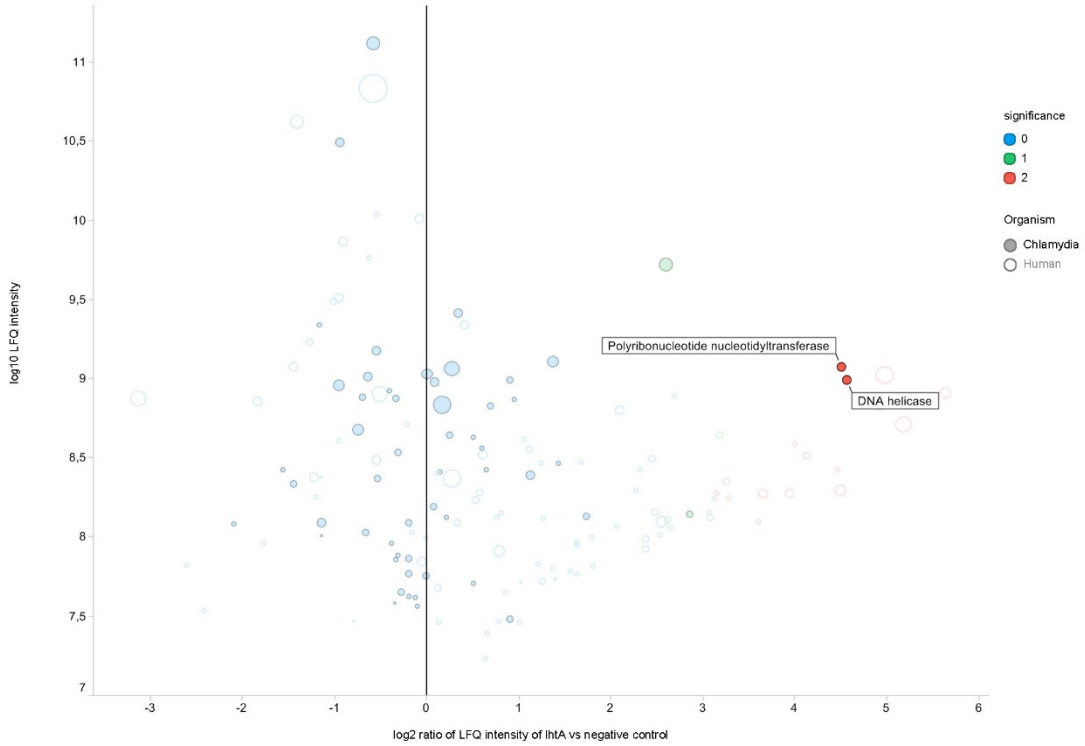


Figure 34: Results of the oligonucleotide pull-down with IhtA compared to the results from the negative control. The Y-axis indicates the intensity of the signal and the X-axis shows the ratio between IhtA and the negative control. The further right a signal is the more it is enriched compared to the negative control. The size of the signals indicates how many razor unique entries were found. The colour indicates the significance of the signal.

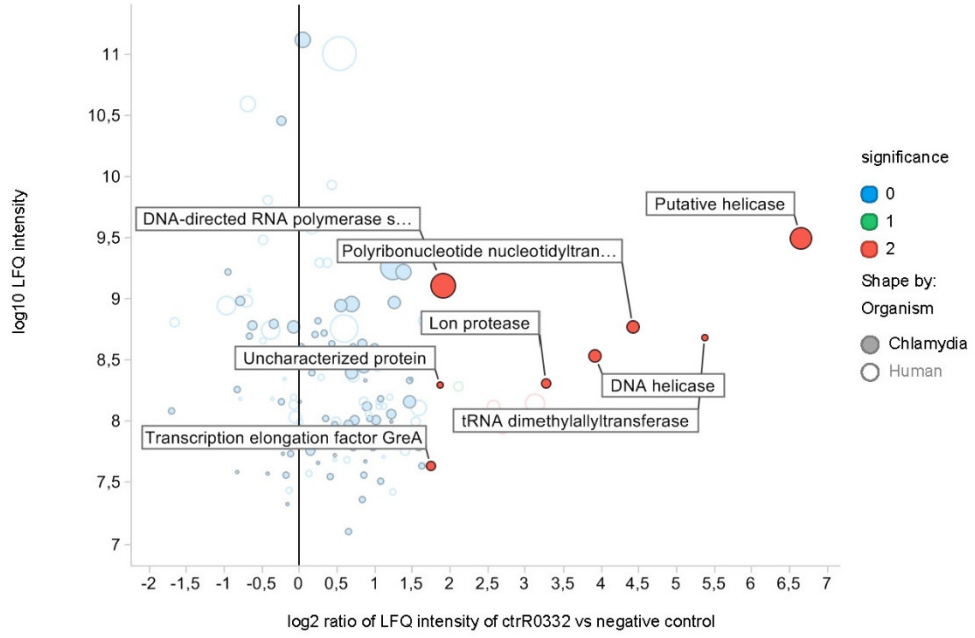


Figure 35: Results of the oligonucleotide pull-down with ctrR0332 compared to the results from the negative control. The Y-axis indicates the intensity of the signal and the X-axis shows the ratio between ctrR0332 and the negative control. The further right a signal is the more it is enriched compared to the negative control. The size of the signals indicates how many razor unique entries were found. The colour indicates the significance of the signal.

The RNA capture pull-down of ctrR0332 showed an interaction with several non-ribosomal proteins including the previously identified protein ctl0077 (Figure 35).

Ctl0077 is a highly specific chlamydial protein. Blast results show that it is highly conserved across *Chlamydia* but cannot be found in other organisms of the RefSeq Database. The protein contains three defined domains: the SWIM domain, a helicase ATP-binding domain and a helicase C-terminal domain. The SWIM domain is a zinc finger-like domain and is predicted to promote DNA binding and protein-protein interactions (Makarova et al., 2002). The protein belongs to the superfamily 1 and 2 helicases and is proposed to be a DEAD/DEAH box protein.

Besides Ctl0077, Ctl0766 was detected using the pull-down. Ctl0766 is annotated as a hypothetical protein. Bao and colleagues showed that ctl0766 interacts with a non-conserved region of σ^{66} and plays a central role in transcription activation. Accordingly, it was renamed as grgA (general regulator of genes A) (Bao et al., 2012).

Additionally, a part of the RNAP β'-subunit was found to interact with ctrR0332 and with the transcriptional elongation factor GreA, which interacts with RNAP. Toulmé and colleagues described that GreA restarts the RNAP when it's stalled during ATC/TAG repetitive sequences (Toulmé et al., 2000). Since RNA was used as a bait in this work instead of DNA, it can be assumed that the discovered interactions are truly RNA-based and represent a novel mode of action.

4. Discussion

4.1. Analysis of TagRNA-seq

Previous studies on *C. trachomatis* and *Chlamydia pneumoniae* utilized 5'P-dependent terminator exonuclease to degrade 5' RNAs, enriching the sample for primary transcripts (Albrecht et al., 2011, Albrecht et al., 2010). This method was first applied in *Helicobacter pylori* (Sharma et al., 2010). In a study by Albrecht and colleagues, 565 TSSs were mapped for *C. pneumoniae* and 363 TSS for *C. trachomatis* (Albrecht et al., 2011, Albrecht et al., 2010). In addition, the authors semi-quantified the transcripts of EB and RB and thus identified ctrR0332 as a highly abundant and differently expressed RNA (20% in RBs and 78% in EBs of all RNAs in *C. trachomatis*) (Albrecht et al., 2010). This work was based on the 454-sequencing technology and only had a low read yield in contrast to the possibilities of the current generation of next-generation sequencers.

In the present work, differential labelling of chlamydial 5' RNA ends was used to differentiate between primary and processed 5' RNA ends. The same method was previously used by Innocenti *et al.* to analyse the transcriptional organizations in *Enterococcus faecalis*. The authors successfully detected 559 TSSs and 352 PSSs (Innocenti et al., 2015). Although *Chlamydia* has a small genome (~1 Mb) it has several intergenic regions creating space for previously undetected transcripts as well as for UTRs which were previously not annotated in the chlamydial genome. In total, 679 TSSs (excluding TSSs that share an identical start with genes from the reference annotation) were detect in this work of which 235 were within genes. In addition, 290 5' UTRs and 154 possible new transcripts in the non-coding regions between genes or on the anti-sense strand of genes were identified. Two of these possible new transcripts were larger versions of the two previously annotated loci infA2 (interferon alpha-2) and tRNA leucin.

Furthermore, it was shown that the well-established RNA IhtA utilizes its own TSS. The transcription from clt0043 upstream of IhtA flows into the sRNA and is possibly further processed, either by eliminating the 3'UTR region or by further increasing the transcriptional levels of IhtA (Figure 6).

The received data for the locus previously annotated as the hypothetical protein CTL0337 and later reannotated by Albrecht and co-workers as ctrR0332 confirm the presence of a shorter transcript (Albrecht et al., 2011, Albrecht et al., 2010). Adding the reads of the PSS, reveals where the processed transcripts give rise to the two shorter forms of the full-length transcript

(Figure 7). Until recently, it was unknown whether this RNA is encoded as a part of the 3' UTR of late transcription unit B *ltuB* or if it has its own transcriptional start site. The TagRNA-seq of the present work shows that this RNA is highly processed. A specific signal for the longer transcript was not detected in northern blots and in the RNA gel. It is conceivable that the long transcript is highly unstable or that it is directly processed, which would make it difficult to detect with the applied methods. The locus of ctrR0332 also has a significant number of primary transcripts although the read number is lower than the number of reads from the PSS library. This would imply the potential presence of a trans-acting sRNA. As already mentioned above, reads covering this locus are highly abundant. Further up- and downstream of the ctrR0332 region, reads in all three libraries were diminished, which further indicates that this RNA is an individual transcript, but highly processed. Another possibility is that this RNA is a 3' UTR-derived sRNA from *ltuB*, which is highly stabilised and a highly abundant transcript.

The motif analyses of the 40 bp upstream sequences of the newly annotated intergenic regions revealed several motifs. The most abundant sequence motif, which was found in 81 sequences upstream of the intergenic annotations, shares similarities to the σ^{66} factor that was also observed by Albrecht and colleagues in their dRNA-seq study (Albrecht et al., 2011, Albrecht et al., 2010). This σ-factor shares high similarities to the σ^{70} factor of *E. coli*. The analysis with tomtom from the MEME toolset searched against known prokaryotic motifs and revealed an association with σ^{70} (rpoD). This motif has the TATAAT -10 box and an A/T rich stretch between the -10 and the -35 box. A high adenine concentration in the -35 box was observed in the sequences used for this analysis. Tan *et al.* postulated that this A/T rich region in the promotor could be used as a non-standard confirmation of the DNA strands and might be recognised by *Chlamydia* RNAP or by another factor or it may play a role in melting of these DNA regions (Tan et al., 1998). The Tomtom analysis further revealed associations with several other motifs of other regulatory functions. Due to the low conservation of chlamydial sequences outside the species, these motifs are probably false positives. Interestingly, the second most abundant motif found in the upstream sequences of newly annotated regions had a high similarity with the mogR (motility gene repressor) binding motif of *Listeria monocytogenes*. In a study by Gründling and co-workers, it was shown that MogR regulates flagellar motility gene expression in a temperature-dependent manner (Grundling et al., 2004). Although *Chlamydia* is generally considered to be non-motile and does not possess flagella, a motility apparatus in marine *Chlamydia* has recently been identified (Collingro et

al., 2017). *Chlamydia* shows restricted export of inclusion membrane protein IncA at 32 °C (Fields et al., 2002) and has a temperature activated HtrA protease which acts both as a chaperone and a protease at 37 °C (Huston et al., 2007). Thus, the newly identified motif in the present work could be either a temperature-dependent promotor sequence or the remainder of the flagella regulon described by Collingro and group (Collingro et al., 2017).

The present results show that a large part of the transcriptome remained uncharacterised in earlier studies. Due to advancements in high-throughput sequencing technology, new regulatory sequences and new previously un-annotated transcripts as well as potentially marginally abundant long ncRNA were identified. In addition, this work provides new knowledge about the organisation of the highly processed transcriptome of *C. trachomatis*.

4.2. Establishment and analysis of the Gradient

In addition to transcriptome profiling, in this study the gradient profile of *C. trachomatis* lysate coupled to high-throughput methods was established and analysed. This procedure coined Grad-seq and was first successfully applied in *Salmonella* by the group around Smirnov. The authors were able to characterize ProQ as a major new sRNA-binding protein in a global approach (Smirnov et al., 2016b). In Grad-seq, a glycerol gradient is used to separate complexes by shape and size (Erickson, 2009, Rederstorff et al., 2010). ProQ or its homolog FinO are known RNA chaperones in *E. coli* (Arthur et al., 2011) and *Legionella pneumophila* (Faucher and Shuman, 2011). Interestingly, the obligate intracellular pathogen *C. trachomatis* does not have any protein homolog to this new RNA chaperon and the previously well characterized RNA chaperons Hfq or CsrA.

To establish Grad-seq for *Chlamydia,* first, conventional gel electrophoresis-based methods were used to find the optimal conditions for this approach. Therefore, several factors had to be adjusted: (a) the input amount of *Chlamydia,* (b) lysis methods the chlamydial pellets were treated with, (c) the conditions for a good reproducibility of the glycerol gradient and (d) finally how to assess the resulting gradients. After switching from a gradient maker device to the gradient master station and increasing the input amount of *Chlamydia* to forty 150 cm^2 cell culture dishes, a reproducible gradient was established that provided enough material to perform the analysis. These two changes and decreasing the lysis time collectively provide optimal conditions to obtain gradients with great reproducibility. These also showed the expected characteristics of ribosomal complexes and sRNAs. Another important step was to move away from protein-based methods, which were characterised by low reproducibility even after keeping conditions stable. The alternative RNA-based approaches were northern blots for identification of specific RNAs and RNA gels to observe the major RNA species in *Chlamydia*.

The optimization of gradient conditions allowed a reliable sample preparation for high-throughput sequencing and MS. The final gradient in this work provided similar results to those obtained by Smirnov and colleagues generated with a Grad-seq study in *Salmonella* (Smirnov et al., 2016b). The references used in the present work behaved similarly as in the study of Smirnov showing peaks for the 30S and 50S subunit and a similar distribution of proteins and RNAs in the RNA-gel (Smirnov et al., 2016b).

In one of the non-ideal gradients, where the lysis was not optimised, the RNAP ββ'-subunit was detected in the same fractions as the sRNA ctrR0332. Albrecht and co-authors hypothesized that ltuB and the sRNA are regulated by σ^{28}. A putative promotor for σ^{28} was identified upstream of ltuB (Albrecht, 2011, Miura et al., 2008). In the work by Albrecht it was observed that the analogue of the *C. pneumoniae* ctrR0332, CPn332, does not co-sediment in the same fractions as this σ-factor σ^{28} (Albrecht, 2011).

Due to the isolation procedure a lot of host contaminants remained in the gradients, which were observed in northern blots with probes specifically designed for the human ribosome. The human material was mostly concentrated in the heavy fractions of the gradient. The chlamydial rRNA was present in the fractions above the human rRNAs, indicating that the used technique not only separates intact chlamydial complexes, but also separated the residual human complexes originating from the host cells.

For the gradient profiling with high-throughput methods, the obtained reads had to be adjusted according to the RNA species in the fractions. Since rRNAs were necessary for quantification and as a reference, the reads for the libraries with predicted high concentration of rRNAs were increased. This effort allowed the generation of enough reads to quantify other RNA species. To counteract the different sequencing depths, an RNA standard was used to calculate the size factors in post-analysis and to adjust the results from the gene quantification.

The RNA sequencing results confirmed the increase of human transcripts in the high glycerol composed fractions. Coverage analysis for each sequencing library revealed that the reads in the light fractions were more homogenous distributed. The higher the glycerol content in the gradient was, the bigger was the size of the complexes. Furthermore, reads were found to accumulate in small regions of the genome, while the rest of the genome had a small coverage. Gene quantification revealed that the highly covered regions were the expected chlamydial rRNAs. The composition of the pellet further supported this fact, since the main RNA species originate from rRNAs and human transcripts were observed. A similar distribution of tRNA and rRNA in a 10 % - 30 % glycerol gradient of human and mouse brain cells was described by the group around Rederstorff (Rederstorff et al., 2010). The authors managed to identify the function of small ncRNAs involved in the ribonucleo-protein particle formation (Rederstorff et al., 2010).

Altogether, the RNAs accumulated in the upper half of the gradient, while proteins were more spread across the gradient, with the majority of the proteins rather accumulating in the upper half of the gradient. An explanation for this observation might be the RNA interactions within *Chlamydia*. It appears that most of chlamydial RNA complexes are rather light compared to protein complexes consisting of several subunits. The exceptions are rRNAs in ribosomal complexes. It has been observed that rRNAs and the respective ribosomal proteins from one ribosomal subunit accumulate in the same gradient fraction confirming that the ribosomal subunit is intact. The sRNA IhtA and the 5S RNA had a similar length but IhtA accumulated in the light/upper fractions of the gradient, indicating no complex formation, while the 5S RNA accumulated in fraction 15, as it is part of large ribosomal complexes. Some of the ribosomal proteins were in the pellet of the gradient indicating the complete 70S ribosome.

The RNAP proteins were also highly abundant in the pellet and peaked within the gradient fractions 5 and 6, while the σ-subunits were present in the very light fractions. The gradient clearly showed that the RNAP holoenzyme is intact and the primary σ-factor σ^{66} is in a complex with the holoenzyme. In contrast, σ^{28} and the anti-σ-factors were not in a complex. It was previously assumed that σ-factors have to dissociate upon transcription elongation (Hsu, 2002). Kapanidis and colleagues showed that σ^{70} can remain in a complex throughout elongation (Kapanidis et al., 2005). The σ^{66}-factor of *C. trachomatis* is related closer to the σ^{70}-factor of *E. coli* than to the chlamydial σ^{28}-factor. The observed complex formation of RNAP with σ^{66} can be explained either by the latter being the major σ-factor (Hua et al., 2009, Douglas and Hatch, 1995, Koehler et al., 1990, Mathews and Stephens, 1999, Nicholson et al., 2003) or σ^{66} is, as described for the σ^{70}, part of the mature RNAP. The lysate did not contain a high percentage of *Chlamydia* differentiating from RB into EB, which could explain why the σ^{28} is not bound to the RNAP, since this differentiation is regulated by the σ^{28}-factor (Brickman et al., 1993, Shen et al., 2004, Yu and Tan, 2003). Furthermore, there was no stress induction during infection, which would be a second circumstance requiring the presence of σ^{28} (Shen et al., 2004, Yu et al., 2006). Although the RsbW of *C. trachomatis* could function as a potential anti-σ^{28}-factor (Hua et al., 2009), no direct inhibitory effect was observed (Hua et al., 2006, Karlinsey and Hughes, 2006). The results show that σ^{28} and the RsbW did not form complexes which further indicates that RsbW is not inhibiting σ^{28}. Additionally, the elusive σ^{54} was detectable in our gradient. This σ-factor is only predicted in chlamydia.

PCA revealed that RNA and proteins were clearly separated due to their above discussed different distribution in the gradient, except for the tRNAs. The *C. trachomatis* bulk of mRNA accumulated in one cluster with the sRNAs being in the fringes of this cluster but spread across a large region. This indicates that there are different kinds of separation characteristics, which in return suggests that the RNA is part of different types of complexes. On the other hand, proteins were separated from the RNAs und were subdivided into two clusters. In one of those clusters the tRNAs were detected whereas the other cluster contained some rRNAs. This pattern can be explained by the separation in the gradient. Ribosomal proteins were observed in the cluster with the rRNAs, revealing these as the only functional complexes visible. The separation in the second principle component reflects the separation of the complexes in the gradient according to their density, explaining the gap between the two protein classes.

PCA of the RNA species including the newly annotated transcripts, showed that the gap between the tRNA cluster and the RNA bulk cluster is filled with several new transcripts.

The newly annotated transcripts from TagRNA-seq were distributed like the known transcripts, which implies that several of them are not in a complex. However, most of them are in a complex suggested by their enrichment in the fractions 6 and 11. In fraction 6 the RNAP was also present, indicating that these RNAs are transcripts. Furthermore, parts of the ribosome were detected in fraction 11 which in return suggests that these RNAs are translated. Taken together, these observations indicate that specific character and function of several transcripts in *C. trachomatis* remain unresolved and further research is required.

The analyses used in this study also confirmed the previous findings by Grieshaber and Tattersall who showed that the sRNA IhtA functions by direct base pairing (Tattersall et al., 2012). The authors postulated that no RNA chaperon is used for the interaction with the RNA of the histone like protein HctB (Tattersall et al., 2012). Ihta is accumulating with the other unbound RNAs in the bulk peak in fraction 2. In this fraction, the RNA of HctB is highly enriched. HctB reaches its highest abundance in fraction 6. The transcript of HctA on the other hand is not accumulating in fraction 2 and, in addition to the peak in fraction 6 with HctB, it is highly enriched in fraction 11. These findings are in line with previous studies showing that IhtA and RNA of HctB are specific binding partners and that RNA of HctA is not a target of sRNA IhtA. This is reflected by the enrichment of IhtA and HctB transcripts in fraction 2, potentially showing RNA-RNA direct base pairing and a decrease of HctB compared to HctA in fraction 11,

where actively translated RNAs were expected (Grieshaber et al., 2006a, Grieshaber et al., 2006b, Grieshaber et al., 2015, Tattersall et al., 2012) (Figure 36).

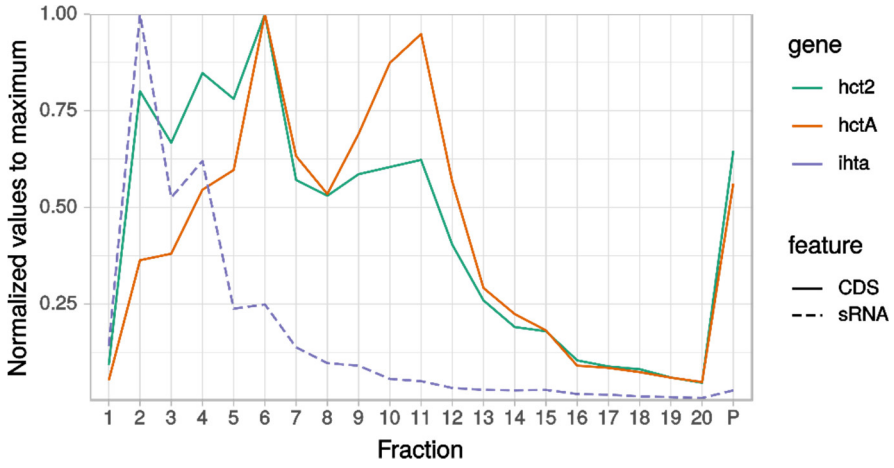


Figure 36: Vizualization of the normalized transcripts across the gradient of the small RNA and the transcripts of hctA and hctB (hct2).

Another transcript that was analysed in-depth is the sRNA ctrR0332, which was observed in the same fraction in which the protein CTL0077 peaked while its transcript was accumulating in the first peak of ctrR0332 (Figure 31B). With the oligo aptamer RNA pull-down, RNA-specific proteins were isolated, and it was validated that the sRNA ctrR0332 is in complex with the protein CTL0077. While it was discovered that IhtA does not have an interaction partner except for its target and the 5S RNA interacted with other rRNAs. For all the other RNA baits the pnp and the UvrD were observed. Both can be part of the degradosome, stress response or repair response (Epshtein et al., 2014, Rosenzweig and Chopra, 2013) and are non-specific binding partners in the pull-down. In addition, the RNAP ββ' and the GreA elongation factor were also pulled down when ctrR0332 was used as a probe for the pull-down. This reflects similar observations which were made for the ncRNA CsrB in *E. coli*. Windbichler and colleagues showed in their studies that several sRNAs are bound by the RNAP β-subunit, while CsrB binds the β'- and α-subunit as well (Windbichler et al., 2008). The protein CTL0077 is highly conserved across *Chlamydia*. It possesses a conserved C-terminal SNF2 region and a N-terminal zinc-finger. In *E. coli* it was previously shown that HepA, a SNF2 protein, associates with the RNAP. In addition, it binds to the RNAP competing over the binding site against the σ^{70} factor (Muzzin et al., 1998). The CTL0077 protein in *Chlamydia* could be part of the regulatory network of the EB to RB transition and vice versa, providing a protein-RNA-chaperon or a protein for the sRNA ctrR0332.

In this study it was shown that Grad-seq can be utilized for intracellular organisms as *C. trachomatis*. Not only previously known complexes were validated, but also a new potential RBP was identified. Taken together, utilizing modern technologies, like TagRNA-seq and Grad-seq in the present study, revealed that the chlamydial transcriptome is even more complex than previous studies suggested.

5. Outlook

In this work it was shown how current technologies and methods can be used to identify new transcripts and to obtain a global interactome of proteins and RNAs. Although a global interactome was characterized, specific RNA-Protein interactions must be individually confirmed. The interactome data presented here give a reference map of potential interaction patterns. One crucial and restricting factor when working with *Chlamydia* is that human or host contaminations cannot be eliminated by the current methods to purify *Chlamydia* in large amounts. To overcome this limitation due to the nature of the pathogen's lifecycle, first steps were made by developing a cell-free (axenic) culture system for *Chlamydia*. By using this, EBs as well as RBs are viable, but they will not replicate nor differentiate (Omsland et al., 2012). Another disadvantage is that common genetic tools used to analyse sRNAs and RBPs are still difficult to implement although most recently, it was shown that the CRISPRi system is feasible in *Chlamydia* (Ouellette, 2018).

Obtaining an antibody for CTL0077, the binding protein of the sRNA ctrR0332, would offer the possibly to perform RIP-seq. Here, target transcripts of the protein are co-immunoprecipitated under native conditions and then sequenced. This results in a library of RNAs bound to the protein of interest (RIP-seq) (Zhao et al., 2010). Other similar methods utilize the T4 RNA ligase, encoded on a plasmid, to link the RNA of interest to the targets within the organism (Gril-seq) (Han et al., 2016). CLIP-seq identifies the targets of the RNA via UV crosslinking of the RNA complexes, trimming RNA regions outside the complex and then proceeding with the extraction of the RBP and the sequencing of the RNAs (Stork and Zheng, 2016). After RNA trimming RNA fragments within the complex could be ligated and extracted as described for CLASH or RIL-seq (Melamed et al., 2016, Helwak and Tollervey, 2016). This procedure additionally shows which RNA is ligated to the target RNA, which is important if the RNP binds multiple sRNAs.

To exceed information about the previously unknown PSS and TSSs obtained in this study, data on termination sites within the organism could provide details of the riboswitches in *C. trachomatis* (Term-seq) (Dar et al., 2016). With the use of third-generation sequencers and unfragmented single-molecule primary transcripts direct profiling of the transcriptome would be possible (Yan et al., 2018). By combining these methods, a complete transcriptome for *C. trachomatis* could be obtained.

Using expansion microscopy, it would be possible to identify the location of proteins and of the sRNAs within *Chlamydia* in order to uncover their mode of operation (Asano et al., 2018). Altogether, newly developed methods for transcriptome analyses, the third-generation sequencers, microscopy and the generation of CRISPRi *Chlamydia* will contribute to major advances in the years to come and thus will help to complete our understanding of the complex organism *C. trachomatis*.

6. References

AHMED, W., ZHENG, K. & LIU, Z. F. 2016. Small Non-Coding RNAs: New Insights in Modulation of Host Immune Response by Intracellular Bacterial Pathogens. *Front Immunol,* 7**,** 431.

AL-YOUNES, H. M., RUDEL, T. & MEYER, T. F. 1999. Characterization and intracellular trafficking pattern of vacuoles containing Chlamydia pneumoniae in human epithelial cells. *Cell Microbiol,* 1**,** 237-47.

ALBRECHT, M. 2011. *Global Transcriptome Analysis of the Human Pathogens Chlamydia trachomatis and Chlamydia pneumoniae.*

ALBRECHT, M., SHARMA, C. M., DITTRICH, M. T., MÜLLER, T., REINHARDT, R., VOGEL, J. & RUDEL, T. 2011. The transcriptional landscape of Chlamydia pneumoniae. *Genome biology,* 12**,** R98-R98.

ALBRECHT, M., SHARMA, C. M., REINHARDT, R., VOGEL, J. & RUDEL, T. 2010. Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome. *Nucleic acids research,* 38**,** 868-877.

ALTUVIA, S., WEINSTEIN-FISCHER, D., ZHANG, A., POSTOW, L. & STORZ, G. 1997. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell,* 90**,** 43-53.

ALTUVIA, S., ZHANG, A., ARGAMAN, L., TIWARI, A. & STORZ, G. 1998. The Escherichia coli OxyS regulatory RNA represses fhlA translation by blocking ribosome binding. *EMBO J,* 17**,** 6069-75.

AMANN, R., SPRINGER, N., SCHONHUBER, W., LUDWIG, W., SCHMID, E. N., MULLER, K. D. & MICHEL, R. 1997. Obligate intracellular bacterial parasites of acanthamoebae related to Chlamydia spp. *Appl Environ Microbiol,* 63**,** 115-21.

ANDREWS, S. 2010. *FastQC: a quality control tool for high throughput sequence data* [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc [Accessed].

ARTHUR, D. C., EDWARDS, R. A., TSUTAKAWA, S., TAINER, J. A., FROST, L. S. & GLOVER, J. N. M. 2011. Mapping interactions between the RNA chaperone FinO and its RNA targets. *Nucleic acids research,* 39**,** 4450-4463.

ASANO, S. M., GAO, R., WASSIE, A. T., TILLBERG, P. W., CHEN, F. & BOYDEN, E. S. 2018. Expansion Microscopy: Protocols for Imaging Proteins and RNA in Cells and Tissues. *Curr Protoc Cell Biol,* 80**,** e56.

BACHMANN, N. L., POLKINGHORNE, A. & TIMMS, P. 2014. Chlamydia genomics: providing novel insights into chlamydial biology. *Trends Microbiol,* 22**,** 464-72.

BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W. & NOBLE, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res,* 37**,** W202-8.

BAO, X., NICKELS, B. E. & FAN, H. 2012. Chlamydia trachomatis protein GrgA activates transcription by contacting the nonconserved region of σ66. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 16870-16875.

BARRY, C. E., 3RD, BRICKMAN, T. J. & HACKSTADT, T. 1993. Hc1-mediated effects on DNA structure: a potential regulator of chlamydial development. *Mol Microbiol,* 9**,** 273-83.

BARRY, C. E., 3RD, HAYES, S. F. & HACKSTADT, T. 1992. Nucleoid condensation in Escherichia coli that express a chlamydial histone homolog. *Science,* 256**,** 377-9.

BAVOIL, P. M., HSIA, R. & OJCIUS, D. M. 2000. Closing in on Chlamydia and its intracellular bag of tricks. *Microbiology,* 146 (Pt 11)**,** 2723-31.

BEATTY, W. L., BELANGER, T. A., DESAI, A. A., MORRISON, R. P. & BYRNE, G. I. 1994. Tryptophan depletion as a mechanism of gamma interferon-mediated chlamydial persistence. *Infect Immun,* 62**,** 3705-11.

BEDSON, S. P. & BLAND, J. O. W. 1932. A Morphological Study of Psittacosis Virus, with the Description of a Developmental Cycle. *British journal of experimental pathology,* 13**,** 461-466.

BELLAND, R. J., ZHONG, G., CRANE, D. D., HOGAN, D., STURDEVANT, D., SHARMA, J., BEATTY, W. L. & CALDWELL, H. D. 2003. Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis. *Proc Natl Acad Sci U S A,* 100**,** 8478-83.

BETTS-HAMPIKIAN, H. J. & FIELDS, K. A. 2010. The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut. *Front Microbiol,* 1**,** 114.

BIRKELUND, S., JOHNSEN, H. & CHRISTIANSEN, G. 1994. Chlamydia trachomatis serovar L2 induces protein tyrosine phosphorylation during uptake by HeLa cells. *Infect Immun,* 62**,** 4900-8.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England),* 30**,** 2114-2120.

BRAIG, K., OTWINOWSKI, Z., HEGDE, R., BOISVERT, D. C., JOACHIMIAK, A., HORWICH, A. L. & SIGLER, P. B. 1994. The crystal structure of the bacterial chaperonln GroEL at 2.8 Å. *Nature,* 371**,** 578-586.

BRANTL, S. 2002. Antisense-RNA regulation and RNA interference. *Biochim Biophys Acta,* 1575**,** 15-25.

BRANTL, S. 2007. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol,* 10**,** 102-9.

BRANTL, S. 2012a. Acting antisense: plasmid- and chromosome-encoded sRNAs from Gram-positive bacteria. *Future Microbiol,* 7**,** 853-71.

BRANTL, S. 2012b. Small Regulatory RNAs (sRNAs): Key Players in Prokaryotic Metabolism, Stress Response, and Virulence. *In:* MALLICK, B. & GHOSH, Z. (eds.) *Regulatory RNAs: Basics, Methods and Applications.* Berlin, Heidelberg: Springer Berlin Heidelberg.

BRENNAN, R. G. & LINK, T. M. 2007. Hfq structure, function and ligand binding. *Curr Opin Microbiol,* 10**,** 125-33.

BRICKMAN, T. J., BARRY, C. E., 3RD & HACKSTADT, T. 1993. Molecular cloning and expression of hctB encoding a strain-variant chlamydial histone-like protein with DNA-binding activity. *J Bacteriol,* 175**,** 4274-81.

BUSH, R. M. & EVERETT, K. D. 2001. Molecular evolution of the Chlamydiaceae. *Int J Syst Evol Microbiol,* 51**,** 203-20.

BUSHNELL, B. 2014. *BBMAP* [Online]. Available: sourceforge.net/projects/bbmap/ [Accessed].

BYRNE, G. I. & OJCIUS, D. M. 2004. Chlamydia and apoptosis: life and death decisions of an intracellular pathogen. *Nat Rev Microbiol,* 2**,** 802-8.

CARABEO, R. 2011. Bacterial subversion of host actin dynamics at the plasma membrane. *Cell Microbiol,* 13**,** 1460-9.

CARABEO, R. A., GRIESHABER, S. S., HASENKRUG, A., DOOLEY, C. & HACKSTADT, T. 2004. Requirement for the Rac GTPase in Chlamydia trachomatis invasion of non-phagocytic cells. *Traffic,* 5**,** 418-25.

CARABEO, R. A. & HACKSTADT, T. 2001. Isolation and characterization of a mutant Chinese hamster ovary cell line that is resistant to Chlamydia trachomatis infection at a novel step in the attachment process. *Infect Immun,* 69**,** 5899-904.

CDC. 2015. *2015 Sexually Transmitted Diseases Treatment Guidelines* [Online]. Available: https://www.cdc.gov/std/tg2015/chlamydia.htm [Accessed 2019].

CHAULK, S. G., SMITH FRIEDAY, M. N., ARTHUR, D. C., CULHAM, D. E., EDWARDS, R. A., SOO, P., FROST, L. S., KEATES, R. A., GLOVER, J. N. & WOOD, J. M. 2011. ProQ is an RNA chaperone that controls ProP levels in Escherichia coli. *Biochemistry,* 50**,** 3095-106.

CLIFTON, D. R., FIELDS, K. A., GRIESHABER, S. S., DOOLEY, C. A., FISCHER, E. R., MEAD, D. J., CARABEO, R. A. & HACKSTADT, T. 2004. A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. *Proc Natl Acad Sci U S A,* 101**,** 10166-71.

COLLINGRO, A., KOSTLBACHER, S., MUSSMANN, M., STEPANAUSKAS, R., HALLAM, S. J. & HORN, M. 2017. Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae. *Isme j,* 11**,** 2334-2344.

COLLINGRO, A., TISCHLER, P., WEINMAIER, T., PENZ, T., HEINZ, E., BRUNHAM, R. C., READ, T. D., BAVOIL, P. M., SACHSE, K., KAHANE, S., FRIEDMAN, M. G., RATTEI, T., MYERS, G. S. & HORN, M. 2011. Unity in variety--the pan-genome of the Chlamydiae. *Mol Biol Evol,* 28**,** 3253-70.

CONSTABLE, F. L. 1959. Psittacosis elementary bodies. *Nature,* 184(Suppl 7)**,** 473-4.

COX, J., HEIN, M. Y., LUBER, C. A., PARON, I., NAGARAJ, N. & MANN, M. 2014. MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction. *Molecular & Cellular Proteomics***,** mcp.M113.031591.

COX, J. & MANN, M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology,* 26**,** 1367.

DAR, D., SHAMIR, M., MELLIN, J. R., KOUTERO, M., STERN-GINOSSAR, N., COSSART, P. & SOREK, R. 2016. Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science,* 352**,** aad9822.

DAUTRY-VARSAT, A., SUBTIL, A. & HACKSTADT, T. 2005. Recent insights into the mechanisms of Chlamydia entry. *Cell Microbiol,* 7**,** 1714-22.

DAVIS, C. H., RAULSTON, J. E. & WYRICK, P. B. 2002. Protein disulfide isomerase, a component of the estrogen receptor complex, is associated with Chlamydia trachomatis serovar E attached to human endometrial epithelial cells. *Infect Immun,* 70**,** 3413-8.

DEKA, S., VANOVER, J., DESSUS-BABUS, S., WHITTIMORE, J., HOWETT, M. K., WYRICK, P. B. & SCHOBORG, R. V. 2006. Chlamydia trachomatis enters a viable but non-cultivable (persistent) state within herpes simplex virus type 2 (HSV-2) co-infected host cells. *Cell Microbiol,* 8**,** 149-62.

DI GIORGIO SILVIA, K. F. 2017. *GRADitude: a computational tool for Grad-seq data analysis* [Online]. Available: https://foerstner-lab.github.io/GRADitude/ [Accessed].

DOMMAN, D. & HORN, M. 2015. Following the Footsteps of Chlamydial Gene Regulation. *Mol Biol Evol,* 32**,** 3035-46.

DOUGLAS, A. L. & HATCH, T. P. 1995. Functional analysis of the major outer membrane protein gene promoters of Chlamydia trachomatis. *Journal of bacteriology,* 177**,** 6286-6289.

DOUGLAS, A. L. & HATCH, T. P. 2000. Expression of the transcripts of the sigma factors and putative sigma factor regulators of Chlamydia trachomatis L2. *Gene,* 247**,** 209-14.

DUSS, O., MICHEL, E., DIARRA DIT KONTE, N., SCHUBERT, M. & ALLAIN, F. H. 2014. Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. *Nucleic Acids Res,* 42**,** 5332-46.

EB, F., ORFILA, J. & LEFEBVRE, J. F. 1976. Ultrastructural study of the development of the agent of ewe's abortion. *J Ultrastruct Res,* 56**,** 177-85.

ELWELL, C., MIRRASHIDI, K. & ENGEL, J. 2016. Chlamydia cell biology and pathogenesis. *Nat Rev Microbiol,* 14**,** 385-400.

ELWELL, C. A., CEESAY, A., KIM, J. H., KALMAN, D. & ENGEL, J. N. 2008. RNA interference screen identifies Abl kinase and PDGFR signaling in Chlamydia trachomatis entry. *PLoS Pathog,* 4**,** e1000021.

ENGEL, J. N., POLLACK, J., MALIK, F. & GANEM, D. 1990. Cloning and characterization of RNA polymerase core subunits of Chlamydia trachomatis by using the polymerase chain reaction. *J Bacteriol,* 172**,** 5732-41.

EPSHTEIN, V., KAMARTHAPU, V., MCGARY, K., SVETLOV, V., UEBERHEIDE, B., PROSHKIN, S., MIRONOV, A. & NUDLER, E. 2014. UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature,* 505**,** 372-7.

ERICKSON, H. P. 2009. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biological procedures online,* 11**,** 32-51.

EVERETT, K. D., BUSH, R. M. & ANDERSEN, A. A. 1999. Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standards for the identification of organisms. *Int J Syst Bacteriol,* 49 Pt 2**,** 415-40.

FAHR, M. J., DOUGLAS, A. L., XIA, W. & HATCH, T. P. 1995. Characterization of late gene promoters of Chlamydia trachomatis. *J Bacteriol,* 177**,** 4252-60.

FANG, F. C. 2005. Sigma cascades in prokaryotic regulatory networks. *Proc Natl Acad Sci U S A,* 102**,** 4933-4.

FAUCHER, S. P. & SHUMAN, H. A. 2011. Small Regulatory RNA and Legionella pneumophila. *Front Microbiol,* 2**,** 98.

FAWAZ, F. S., VAN OOIJ, C., HOMOLA, E., MUTKA, S. C. & ENGEL, J. N. 1997. Infection with Chlamydia trachomatis alters the tyrosine phosphorylation and/or localization of several host cell proteins including cortactin. *Infect Immun,* 65**,** 5301-8.

FIELDS, K. A., FISCHER, E. & HACKSTADT, T. 2002. Inhibition of fusion of Chlamydia trachomatis inclusions at 32 degrees C correlates with restricted export of IncA. *Infect Immun,* 70**,** 3816-23.

FÖRSTNER, K. U., VOGEL, J. & SHARMA, C. M. 2014. READemption—a tool for the computational analysis of deep-sequencing–based transcriptome data. *Bioinformatics,* 30**,** 3421-3423.

FRANZE DE FERNANDEZ, M. T., EOYANG, L. & AUGUST, J. T. 1968. Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature,* 219**,** 588-90.

FREESE, N. H., NORRIS, D. C. & LORAINE, A. E. 2016. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics (Oxford, England),* 32**,** 2089-2095.

FUKUSHI, H. & HIRAI, K. 1992. Proposal of Chlamydia pecorum sp. nov. for Chlamydia strains derived from ruminants. *Int J Syst Bacteriol,* 42**,** 306-8.

GAYLORD, W. H., JR. 1954. Intracellular forms of meningopneumonitis virus. *J Exp Med,* 100**,** 575-80.

GORDON, F. B. & QUAN, A. L. 1965. Occurence of Glycogen in Inclusions of the Psittacosis-Lymphogranuloma Venereum-Trachoma Agents. *J Infect Dis,* 115**,** 186-96.

GOTTESMAN, S. 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet,* 21**,** 399-404.

GOTTESMAN, S. & STORZ, G. 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol,* 3.

GRAYSTON, J. T., KUO, C. C., WANG, S. P. & ALTMAN, J. 1986. A new Chlamydia psittaci strain, TWAR, isolated in acute respiratory tract infections. *N Engl J Med,* 315**,** 161-8.

GRIESHABER, N. A., GRIESHABER, S. S., FISCHER, E. R. & HACKSTADT, T. 2006a. A small RNA inhibits translation of the histone-like protein Hc1 in Chlamydia trachomatis. *Mol Microbiol,* 59**,** 541-50.

GRIESHABER, N. A., SAGER, J. B., DOOLEY, C. A., HAYES, S. F. & HACKSTADT, T. 2006b. Regulation of the Chlamydia trachomatis histone H1-like protein Hc2 is IspE dependent and IhtA independent. *J Bacteriol,* 188**,** 5289-92.

GRIESHABER, N. A., TATTERSALL, J. S., LIGUORI, J., LIPAT, J. N., RUNAC, J. & GRIESHABER, S. S. 2015. Identification of the base-pairing requirements for repression of hctA translation by the small RNA IhtA leads to the discovery of a new mRNA target in Chlamydia trachomatis. *PLoS One,* 10**,** e0116593.

GRIMWOOD, J., OLINGER, L. & STEPHENS, R. S. 2001. Expression of Chlamydia pneumoniae polymorphic membrane protein family genes. *Infect Immun,* 69**,** 2383-9.

GRUNDLING, A., BURRACK, L. S., BOUWER, H. G. & HIGGINS, D. E. 2004. Listeria monocytogenes regulates flagellar motility gene expression through MogR, a transcriptional repressor required for virulence. *Proc Natl Acad Sci U S A,* 101**,** 12318-23.

GU, L., WENMAN, W. M., REMACHA, M., MEUSER, R., COFFIN, J. & KAUL, R. 1995. Chlamydia trachomatis RNA polymerase alpha subunit: sequence and structural analysis. *J Bacteriol,* 177**,** 2594-601.

GUBBINS, M. J., ARTHUR, D. C., GHETU, A. F., GLOVER, J. N. & FROST, L. S. 2003. Characterizing the structural features of RNA/RNA interactions of the F-plasmid FinOP fertility inhibition system. *J Biol Chem,* 278**,** 27663-71.

HACKSTADT, T. 2000. Redirection of host vesicle trafficking pathways by intracellular parasites. *Traffic,* 1**,** 93-9.

HACKSTADT, T., BAEHR, W. & YING, Y. 1991. Chlamydia trachomatis developmentally regulated protein is homologous to eukaryotic histone H1. *Proc Natl Acad Sci U S A,* 88**,** 3937-41.

HACKSTADT, T., FISCHER, E. R., SCIDMORE, M. A., ROCKEY, D. D. & HEINZEN, R. A. 1997. Origins and functions of the chlamydial inclusion. *Trends Microbiol,* 5**,** 288-93.

HAIDER, S., WAGNER, M., SCHMID, M. C., SIXT, B. S., CHRISTIAN, J. G., HACKER, G., PICHLER, P., MECHTLER, K., MULLER, A., BARANYI, C., TOENSHOFF, E. R., MONTANARO, J. & HORN, M. 2010. Raman microspectroscopy reveals long-term extracellular activity of Chlamydiae. *Mol Microbiol,* 77**,** 687-700.

HALBERSTÄDTER, L. & PROWAZEK, S. V. 1907. Zur Atiologie des Trachoms. *Deutsche Medizinische Wochenschrift.*

HALDENWANG, W. G. 1995. The sigma factors of Bacillus subtilis. *Microbiol Rev,* 59**,** 1-30.

HAN, K., TJADEN, B. & LORY, S. 2016. GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. *Nature Microbiology,* 2**,** 16239.

HELWAK, A. & TOLLERVEY, D. 2016. Identification of miRNA-Target RNA Interactions Using CLASH. *Methods Mol Biol,* 1358**,** 229-51.

HODINKA, R. L., DAVIS, C. H., CHOONG, J. & WYRICK, P. B. 1988. Ultrastructural study of endocytosis of Chlamydia trachomatis by McCoy cells. *Infect Immun,* 56**,** 1456-63.

HOFFMANN, S., OTTO, C., KURTZ, S., SHARMA, C. M., KHAITOVICH, P., VOGEL, J., STADLER, P. F. & HACKERMÜLLER, J. 2009. Fast Mapping of Short Sequences with Mismatches, Insertions and Deletions Using Index Structures. *PLOS Computational Biology,* 5**,** e1000502.

HORN, M., COLLINGRO, A., SCHMITZ-ESSER, S., BEIER, C. L., PURKHOLD, U., FARTMANN, B., BRANDT, P., NYAKATURA, G. J., DROEGE, M., FRISHMAN, D., RATTEI, T., MEWES, H. W. & WAGNER, M. 2004. Illuminating the evolutionary history of chlamydiae. *Science,* 304**,** 728-30.

HSU, L. M. 2002. Promoter clearance and escape in prokaryotes. *Biochim Biophys Acta,* 1577**,** 191-207.

HUA, L., HEFTY, P. S., LEE, Y. J., LEE, Y. M., STEPHENS, R. S. & PRICE, C. W. 2006. Core of the partner switching signalling mechanism is conserved in the obligate intracellular pathogen Chlamydia trachomatis. *Mol Microbiol,* 59**,** 623-36.

HUA, Z., RAO, X., FENG, X., LUO, X., LIANG, Y. & SHEN, L. 2009. Mutagenesis of Region 4 of Sigma 28 from Chlamydia trachomatis Defines Determinants for Protein-Protein and Protein-DNA Interactions. *Journal of Bacteriology,* 191**,** 651.

HUSTON, W. M., SWEDBERG, J. E., HARRIS, J. M., WALSH, T. P., MATHEWS, S. A. & TIMMS, P. 2007. The temperature activated HtrA protease from pathogen Chlamydia trachomatis acts as both a chaperone and protease at 37 degrees C. *FEBS Lett,* 581**,** 3382-6.

INNOCENTI, N., GOLUMBEANU, M., FOUQUIER D'HEROUEL, A., LACOUX, C., BONNIN, R. A., KENNEDY, S. P., WESSNER, F., SERROR, P., BOULOC, P., REPOILA, F. & AURELL, E. 2015. Whole-genome mapping of 5' RNA ends in bacteria by tagged sequencing: a comprehensive view in Enterococcus faecalis. *Rna,* 21**,** 1018-30.

JACKSON, D. W., SUZUKI, K., OAKFORD, L., SIMECKA, J. W., HART, M. E. & ROMEO, T. 2002. Biofilm formation and dispersal under the influence of the global regulator CsrA of Escherichia coli. *J Bacteriol,* 184**,** 290-301.

JEROME, L. J., VAN BIESEN, T. & FROST, L. S. 1999. Degradation of FinP antisense RNA from F-like plasmids: the RNA-binding protein, FinO, protects FinP from ribonuclease E. *J Mol Biol,* 285**,** 1457-73.

JEWETT, T. J., FISCHER, E. R., MEAD, D. J. & HACKSTADT, T. 2006. Chlamydial TARP is a bacterial nucleator of actin. *Proc Natl Acad Sci U S A,* 103**,** 15599-604.

JUTRAS, I., ABRAMI, L. & DAUTRY-VARSAT, A. 2003. Entry of the lymphogranuloma venereum strain of Chlamydia trachomatis into host cells involves cholesterol-rich membrane domains. *Infect Immun,* 71**,** 260-6.

KAHANE, S., METZER, E. & FRIEDMAN, M. G. 1995. Evidence that the novel microorganism 'Z' may belong to a new genus in the family Chlamydiaceae. *FEMS Microbiol Lett,* 126**,** 203-7.

KALMAN, S., MITCHELL, W., MARATHE, R., LAMMEL, C., FAN, J., HYMAN, R. W., OLINGER, L., GRIMWOOD, J., DAVIS, R. W. & STEPHENS, R. S. 1999. Comparative genomes of Chlamydia pneumoniae and C. trachomatis. *Nat Genet,* 21**,** 385-9.

KAPANIDIS, A. N., MARGEAT, E., LAURENCE, T. A., DOOSE, S., HO, S. O., MUKHOPADHYAY, J., KORTKHONJIA, E., MEKLER, V., EBRIGHT, R. H. & WEISS, S. 2005. Retention of transcription initiation factor sigma70 in transcription elongation: single-molecule analysis. *Molecular cell,* 20**,** 347-356.

KARLINSEY, J. E. & HUGHES, K. T. 2006. Genetic transplantation: Salmonella enterica serovar Typhimurium as a host to study sigma factor and anti-sigma factor interactions in genetically intractable systems. *J Bacteriol,* 188**,** 103-14.

KAZMIERCZAK, M. J., WIEDMANN, M. & BOOR, K. J. 2005. Alternative sigma factors and their roles in bacterial virulence. *Microbiol Mol Biol Rev,* 69**,** 527-43.

KERN, J. M., MAASS, V. & MAASS, M. 2009. Molecular pathogenesis of chronic Chlamydia pneumoniae infection: a brief overview. *Clin Microbiol Infect,* 15**,** 36-41.

KIM, J. H., JIANG, S., ELWELL, C. A. & ENGEL, J. N. 2011. Chlamydia trachomatis co-opts the FGF2 signaling pathway to enhance infection. *PLoS Pathog,* 7**,** e1002285.

KOEHLER, J. E., BURGESS, R. R., THOMPSON, N. E. & STEPHENS, R. S. 1990. Chlamydia trachomatis RNA polymerase major sigma subunit. Sequence and structural comparison of conserved and unique regions with Escherichia coli sigma 70 and Bacillus subtilis sigma 43. *J Biol Chem,* 265**,** 13206-14.

KOONIN, E. V. & WOLF, Y. I. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res,* 36**,** 6688-719.

KORAIMANN, G., TEFERLE, K., MARKOLIN, G., WOGER, W. & HOGENAUER, G. 1996. The FinOP repressor system of plasmid R1: analysis of the antisense RNA control of traJ expression and conjugative DNA transfer. *Mol Microbiol,* 21**,** 811-21.

KUNTE, H. J., CRANE, R. A., CULHAM, D. E., RICHMOND, D. & WOOD, J. M. 1999. Protein ProQ influences osmotic activation of compatible solute transporter ProP in Escherichia coli K-12. *J Bacteriol,* 181**,** 1537-43.

KUO, C., WANG, S. & GRAYSTON, J. T. 1972. Differentiation of TRIC and LGV organisms based on enhancement of infectivity by DEAE-dextran in cell culture. *J Infect Dis,* 125**,** 313-7.

KUO, C. C., PUOLAKKAINEN, M., LIN, T. M., WITTE, M. & CAMPBELL, L. A. 2002. Mannose-receptor positive and negative mouse macrophages differ in their susceptibility to infection by Chlamydia species. *Microb Pathog,* 32**,** 43-8.

KUO, C. C., SHOR, A., CAMPBELL, L. A., FUKUSHI, H., PATTON, D. L. & GRAYSTON, J. T. 1993. Demonstration of Chlamydia pneumoniae in atherosclerotic lesions of coronary arteries. *J Infect Dis,* 167**,** 841-9.

KUO, C. C., WANG, S. P. & GRAYSTON, J. T. 1973. Effect of polycations, polyanions and neuraminidase on the infectivity of trachoma-inclusin conjunctivitis and lymphogranuloma venereum organisms HeLa cells: sialic acid residues as possible receptors for trachoma-inclusion conjunction. *Infect Immun,* 8**,** 74-9.

LENZ, D. H., MILLER, M. B., ZHU, J., KULKARNI, R. V. & BASSLER, B. L. 2005. CsrA and three redundant small RNAs regulate quorum sensing in Vibrio cholerae. *Mol Microbiol,* 58**,** 1186-202.

LIN, H. S. & MOULDER, J. W. 1966. Patterns of response to sulfadiazine, D-cycloserine and D-alanine in members of the psittacosis group. *J Infect Dis,* 116**,** 372-6.

LINK, T. M., VALENTIN-HANSEN, P. & BRENNAN, R. G. 2009. Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci U S A,* 106**,** 19292-7.

LONETTO, M., GRIBSKOV, M. & GROSS, C. A. 1992. The sigma 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology,* 174**,** 3843.

MAJDALANI, N., VANDERPOOL, C. K. & GOTTESMAN, S. 2005. Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol,* 40**,** 93-113.

MAKAROVA, K. S., ARAVIND, L. & KOONIN, E. V. 2002. SWIM, a novel Zn-chelating domain present in bacteria, archaea and eukaryotes. *Trends Biochem Sci,* 27**,** 384-6.

MALHOTRA, M., SOOD, S., MUKHERJEE, A., MURALIDHAR, S. & BALA, M. 2013. Genital Chlamydia trachomatis: an update. *Indian J Med Res,* 138**,** 303-16.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal,* 17**,** 3.

MASSE, E., ESCORCIA, F. E. & GOTTESMAN, S. 2003. Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes Dev,* 17**,** 2374-83.

MASSE, E. & GOTTESMAN, S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci U S A,* 99**,** 4620-5.

MATHEWS, S. A., DOUGLAS, A., SRIPRAKASH, K. S. & HATCH, T. P. 1993. In vitro transcription in Chlamydia psittaci and Chlamydia trachomatis. *Mol Microbiol,* 7**,** 937-46.

MATHEWS, S. A. & STEPHENS, R. S. 1999. DNA structure and novel amino and carboxyl termini of the Chlamydia sigma 70 analogue modulate promoter recognition. *Microbiology,* 145 (Pt 7)**,** 1671-81.

MAURER, A. P., MEHLITZ, A., MOLLENKOPF, H. J. & MEYER, T. F. 2007. Gene expression profiles of Chlamydophila pneumoniae during the developmental cycle and iron depletion-mediated persistence. *PLoS Pathog,* 3**,** e83.

MCCLARTY, G. 1994. Chlamydiae and the biochemistry of intracellular parasitism. *Trends Microbiol,* 2**,** 157-64.

MELAMED, S., PEER, A., FAIGENBAUM-ROMM, R., GATT, Y. E., REISS, N., BAR, A., ALTUVIA, Y., ARGAMAN, L. & MARGALIT, H. 2016. Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol Cell,* 63**,** 884-97.

MERRICK, M. J. 1993. In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N). *Mol Microbiol,* 10**,** 903-9.

MIURA, K., TOH, H., HIRAKAWA, H., SUGII, M., MURATA, M., NAKAI, K., TASHIRO, K., KUHARA, S., AZUMA, Y. & SHIRAI, M. 2008. Genome-wide analysis of Chlamydophila pneumoniae gene expression at the late stage of infection. *DNA research : an international journal for rapid publication of reports on genes and genomes,* 15**,** 83-91.

MORAN, N. A. 2002. Microbial Minimalism: Genome Reduction in Bacterial Pathogens. *Cell,* 108**,** 583-586.

MOULDER, J. W. 1966. The relation of the psittacosis group (Chlamydiae) to bacteria and viruses. *Annu Rev Microbiol,* 20**,** 107-30.

MOULDER, J. W. 1991. Interaction of chlamydiae and host cells in vitro. *Microbiol Rev,* 55**,** 143-90.

MOULDER, J. W., LEVY, N. J. & SCHULMAN, L. P. 1980. Persistent infection of mouse fibroblasts (L cells) with Chlamydia psittaci: evidence for a cryptic chlamydial form. *Infect Immun,* 30**,** 874-83.

MUZZIN, O., CAMPBELL, E. A., XIA, L., SEVERINOVA, E., DARST, S. A. & SEVERINOV, K. 1998. Disruption of Escherichia coli HepA, an RNA Polymerase-associated Protein, Causes UV Sensitivity. *Journal of Biological Chemistry,* 273**,** 15157-15161.

NICHOLSON, T. L., OLINGER, L., CHONG, K., SCHOOLNIK, G. & STEPHENS, R. S. 2003. Global stage-specific gene regulation during the developmental cycle of Chlamydia trachomatis. *Journal of bacteriology,* 185**,** 3179-3189.

NUNES, A. & GOMES, J. P. 2014. Evolution, phylogeny, and molecular epidemiology of Chlamydia. *Infect Genet Evol,* 23**,** 49-64.

OLEJNICZAK, M. 2011. Despite similar binding to the Hfq protein regulatory RNAs widely differ in their competition performance. *Biochemistry,* 50**,** 4427-40.

OMSLAND, A., SAGER, J., NAIR, V., STURDEVANT, D. E. & HACKSTADT, T. 2012. Developmental stage-specific metabolic and transcriptional activity of Chlamydia trachomatis in an axenic medium. *Proceedings of the National Academy of Sciences,* 109**,** 19781.

OUELLETTE, S. P. 2018. Feasibility of a Conditional Knockout System for Chlamydia Based on CRISPR Interference. *Front Cell Infect Microbiol,* 8**,** 59.

PAGET, M. S. & HELMANN, J. D. 2003. The sigma70 family of sigma factors. *Genome Biol,* 4**,** 203.

PEDERSEN, L. B., BIRKELUND, S. & CHRISTIANSEN, G. 1994. Interaction of the Chlamydia trachomatis histone H1-like protein (Hc1) with DNA and RNA causes repression of transcription and translation in vitro. *Mol Microbiol,* 11**,** 1085-98.

PEDERSEN, L. B., BIRKELUND, S. & CHRISTIANSEN, G. 1996. Purification of recombinant Chlamydia trachomatis histone H1-like protein Hc2, and comparative functional analysis of Hc2 and Hc1. *Mol Microbiol,* 20**,** 295-311.

PERARA, E., GANEM, D. & ENGEL, J. N. 1992. A developmentally regulated chlamydial gene with apparent homology to eukaryotic histone H1. *Proc Natl Acad Sci U S A,* 89**,** 2125-9.

PEREZ-REYTOR, D., PLAZA, N., ESPEJO, R. T., NAVARRETE, P., BASTIAS, R. & GARCIA, K. 2016. Role of Non-coding Regulatory RNA in the Virulence of Human Pathogenic Vibrios. *Front Microbiol,* 7**,** 2160.

PLAUNT, M. R. & HATCH, T. P. 1988. Protein synthesis early in the developmental cycle of Chlamydia psittaci. *Infect Immun,* 56**,** 3021-5.

PRICE, C. W. 2002. Bacillus subtilis and Its Closest Relatives: from Genes to Cells. American Society of Microbiology.

PUOLAKKAINEN, M., KUO, C. C. & CAMPBELL, L. A. 2005. Chlamydia pneumoniae uses the mannose 6-phosphate/insulin-like growth factor 2 receptor for infection of endothelial cells. *Infect Immun,* 73**,** 4620-5.

QAMRA, R., SRINIVAS, V. & MANDE, S. C. 2004. Mycobacterium tuberculosis GroEL Homologues Unusually Exist as Lower Oligomers and Retain the Ability to Suppress Aggregation of Substrate Proteins. *Journal of Molecular Biology,* 342**,** 605-617.

RAO, X., DEIGHAN, P., HUA, Z., HU, X., WANG, J., LUO, M., WANG, J., LIANG, Y., ZHONG, G., HOCHSCHILD, A. & SHEN, L. 2009a. A regulator from Chlamydia trachomatis modulates the activity of RNA polymerase through direct interaction with the beta subunit and the primary sigma subunit. *Genes Dev,* 23**,** 1818-29.

RAO, X., DEIGHAN, P., HUA, Z., HU, X., WANG, J., LUO, M., WANG, J., LIANG, Y., ZHONG, G., HOCHSCHILD, A. & SHEN, L. 2009b. A regulator from Chlamydia trachomatis modulates the activity of RNA polymerase through direct interaction with the β subunit and the primary σ subunit. *Genes & Development,* 23**,** 1818-1829.

RAPPSILBER, J., ISHIHAMA, Y. & 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**,** 663-70.

READ, T. D., BRUNHAM, R. C., SHEN, C., GILL, S. R., HEIDELBERG, J. F., WHITE, O., HICKEY, E. K., PETERSON, J., UTTERBACK, T., BERRY, K., BASS, S., LINHER, K., WEIDMAN, J., KHOURI, H., CRAVEN, B., BOWMAN, C., DODSON, R., GWINN, M., NELSON, W., DEBOY, R., KOLONAY, J., MCCLARTY, G., SALZBERG, S. L., EISEN, J. & FRASER, C. M. 2000. Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. *Nucleic Acids Res,* 28**,** 1397-406.

REDERSTORFF, M., BERNHART, S. H., TANZER, A., ZYWICKI, M., PERFLER, K., LUKASSER, M., HOFACKER, I. L. & HÜTTENHOFER, A. 2010. RNPomics: defining the ncRNA transcriptome by cDNA library generation from ribonucleo-protein particles. *Nucleic acids research,* 38**,** e113-e113.

REPOILA, F. & DARFEUILLE, F. 2009. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biol Cell,* 101**,** 117-31.

REPOILA, F. & GOTTESMAN, S. 2003. Temperature sensing by the dsrA promoter. *J Bacteriol,* 185**,** 6609-14.

RKI. 2010. *Chlamydiosen (Teil 1): Erkrankungen durch Chlamydia trachomatis* [Online]. Available: https://www.rki.de/DE/Content/Infekt/EpidBull/Merkblaetter/Ratgeber_Chlamydiosen_Teil1.html [Accessed 2019].

ROSARIO, C. J. & TAN, M. 2012. The early gene product EUO is a transcriptional repressor that selectively regulates promoters of Chlamydia late genes. *Mol Microbiol,* 84**,** 1097-107.

ROSENZWEIG, J. A. & CHOPRA, A. K. 2013. The exoribonuclease Polynucleotide Phosphorylase influences the virulence and stress responses of yersiniae and many other pathogens. *Frontiers in cellular and infection microbiology,* 3**,** 81-81.

SABNIS, N. A., YANG, H. & ROMEO, T. 1995. Pleiotropic regulation of central carbohydrate metabolism in Escherichia coli via the gene csrA. *J Biol Chem,* 270**,** 29096-104.

SACHSE, K., BAVOIL, P. M., KALTENBOECK, B., STEPHENS, R. S., KUO, C. C., ROSSELLO-MORA, R. & HORN, M. 2015. Emendation of the family Chlamydiaceae: proposal of a single genus, Chlamydia, to include all currently recognized species. *Syst Appl Microbiol,* 38**,** 99-103.

SACHSE, K., LAROUCAU, K., RIEGE, K., WEHNER, S., DILCHER, M., CREASY, H. H., WEIDMANN, M., MYERS, G., VORIMORE, F., VICARI, N., MAGNINO, S., LIEBLER-TENORIO, E., RUETTGER, A., BAVOIL, P. M., HUFERT, F. T., ROSSELLO-MORA, R. & MARZ, M. 2014. Evidence for the existence of two new members of the family Chlamydiaceae and proposal of Chlamydia avium sp. nov. and Chlamydia gallinacea sp. nov. *Syst Appl Microbiol,* 37**,** 79-88.

SAUER, E., SCHMIDT, S. & WEICHENRIEDER, O. 2012. Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc Natl Acad Sci U S A,* 109**,** 9396-401.

SCHUMACHER, M. A., PEARSON, R. F., MOLLER, T., VALENTIN-HANSEN, P. & BRENNAN, R. G. 2002. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J,* 21**,** 3546-56.

SCIDMORE-CARLSON, M. & HACKSTADT, T. 2000. Chlamydia internalization and intracellular fate. *Subcell Biochem,* 33**,** 459-78.

SHARMA, C. M., HOFFMANN, S., DARFEUILLE, F., REIGNIER, J., FINDEISS, S., SITTKA, A., CHABAS, S., REICHE, K., HACKERMULLER, J., REINHARDT, R., STADLER, P. F. & VOGEL, J. 2010. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature,* 464**,** 250-5.

SHAW, A. C., GEVAERT, K., DEMOL, H., HOORELBEKE, B., VANDEKERCKHOVE, J., LARSEN, M. R., ROEPSTORFF, P., HOLM, A., CHRISTIANSEN, G. & BIRKELUND, S. 2002. Comparative proteome analysis of Chlamydia trachomatis serovar A, D and L2. *Proteomics,* 2**,** 164-86.

SHEMER, Y. & SAROV, I. 1985. Inhibition of growth of Chlamydia trachomatis by human gamma interferon. *Infect Immun,* 48**,** 592-6.

SHEN, L., LI, M. & ZHANG, Y. X. 2004. Chlamydia trachomatis sigma28 recognizes the fliC promoter of Escherichia coli and responds to heat shock in chlamydiae. *Microbiology,* 150**,** 205-15.

SHIRAI, M., HIRAKAWA, H., KIMOTO, M., TABUCHI, M., KISHI, F., OUCHI, K., SHIBA, T., ISHII, K., HATTORI, M., KUHARA, S. & NAKAZAWA, T. 2000. Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA. *Nucleic Acids Res,* 28**,** 2311-4.

SIGALOVA, O., CHAPLIN, A. V., BOCHKAREVA, O. O., SHELYAKIN, P. V., FILARETOV, V. A., AKKURATOV, E. E., BURSKAYA, V. & GELFAND, M. S. 2019. Chlamydia pan-genomic

analysis reveals balance between host adaptation and selective pressure to genome reduction. *bioRxiv*, 506121.

SIXT, B. S., HEINZ, C., PICHLER, P., HEINZ, E., MONTANARO, J., OP DEN CAMP, H. J., AMMERER, G., MECHTLER, K., WAGNER, M. & HORN, M. 2011. Proteomic analysis reveals a virtually complete set of proteins for translation and energy generation in elementary bodies of the amoeba symbiont Protochlamydia amoebophila. *Proteomics,* 11**,** 1868-92.

SKILTON, R. J., CUTCLIFFEN, L. T., BARLOW, D., WANG, Y., SALIM, O., LAMBDEN, P. R. & CLARKE, I. N. 2009. Penicillin induced persistence in Chlamydia trachomatis: high quality time lapse video analysis of the developmental cycle. *PLoS One,* 4**,** e7723.

SKIPP, P., ROBINSON, J., O'CONNOR, C. D. & CLARKE, I. N. 2005. Shotgun proteomic analysis of Chlamydia trachomatis. *Proteomics,* 5**,** 1558-73.

SMIRNOV, A., FORSTNER, K. U., HOLMQVIST, E., OTTO, A., GUNSTER, R., BECHER, D., REINHARDT, R. & VOGEL, J. 2016a. 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., FÖRSTNER, K. U., HOLMQVIST, E., OTTO, A., GÜNSTER, R., BECHER, D., REINHARDT, R. & VOGEL, J. 2016b. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proceedings of the National Academy of Sciences of the United States of America,* 113**,** 11591-11596.

SMIRNOV, A., WANG, C., DREWRY, L. L. & VOGEL, J. 2017. Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. *EMBO J,* 36**,** 1029-1045.

SOREK, R., KUNIN, V. & HUGENHOLTZ, P. 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol,* 6**,** 181-6.

SPAETH, K. E., CHEN, Y. S. & VALDIVIA, R. H. 2009. The Chlamydia type III secretion system C-ring engages a chaperone-effector protein complex. *PLoS Pathog,* 5**,** e1000579.

STEPHENS, R. S., KALMAN, S., LAMMEL, C., FAN, J., MARATHE, R., ARAVIND, L., MITCHELL, W., OLINGER, L., TATUSOV, R. L., ZHAO, Q., KOONIN, E. V. & DAVIS, R. W. 1998. Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. *Science,* 282**,** 754-9.

STOJANOV, M., BAUD, D., GREUB, G. & VULLIEMOZ, N. 2018. Male infertility: the intracellular bacterial hypothesis. *New Microbes New Infect,* 26**,** 37-41.

STORK, C. & ZHENG, S. 2016. Genome-Wide Profiling of RNA-Protein Interactions Using CLIP-Seq. *Methods Mol Biol,* 1421**,** 137-51.

STUART, E. S., WEBLEY, W. C. & NORKIN, L. C. 2003. Lipid rafts, caveolae, caveolin-1, and entry by Chlamydiae into host cells. *Exp Cell Res,* 287**,** 67-78.

SU, H., RAYMOND, L., ROCKEY, D. D., FISCHER, E., HACKSTADT, T. & CALDWELL, H. D. 1996. A recombinant Chlamydia trachomatis major outer membrane protein binds to heparan sulfate receptors on epithelial cells. *Proc Natl Acad Sci U S A,* 93**,** 11143-8.

SUBTIL, A., BLOCKER, A. & DAUTRY-VARSAT, A. 2000. Type III secretion system in Chlamydia species: identified members and candidates. *Microbes Infect,* 2**,** 367-9.

SWANSON, A. F. & KUO, C. C. 1991. Evidence that the major outer membrane protein of Chlamydia trachomatis is glycosylated. *Infect Immun,* 59**,** 2120-5.

TAN, M. & BAVOIL, P. 2012. *Intracellular Pathogens I: Chlamydiales*, American Society of Microbiology.

TAN, M., GAAL, T., GOURSE, R. L. & ENGEL, J. N. 1998. Mutational analysis of the Chlamydia trachomatis rRNA P1 promoter defines four regions important for transcription in vitro. *J Bacteriol,* 180**,** 2359-66.

TATTERSALL, J., RAO, G. V., RUNAC, J., HACKSTADT, T., GRIESHABER, S. S. & GRIESHABER, N. A. 2012. Translation inhibition of the developmental cycle protein HctA by the small RNA IhtA is conserved across Chlamydia. *PLoS One,* 7**,** e47439.

TOULMÉ, F., MOSRIN-HUAMAN, C., SPARKOWSKI, J., DAS, A., LENG, M. & RAHMOUNI, A. R. 2000. GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *The EMBO journal,* 19**,** 6853-6859.

TREVORS, J. T. 1996. Genome size in bacteria. *Antonie van Leeuwenhoek,* 69**,** 293-303.

UPDEGROVE, T. B., SHABALINA, S. A. & STORZ, G. 2015. How do base-pairing small RNAs evolve? *FEMS Microbiol Rev,* 39**,** 379-91.

VAN ASSCHE, E., VAN PUYVELDE, S., VANDERLEYDEN, J. & STEENACKERS, H. P. 2015. RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Front Microbiol,* 6**,** 141.

VANOVER, J., SUN, J., DEKA, S., KINTNER, J., DUFFOURC, M. M. & SCHOBORG, R. V. 2008. Herpes simplex virus co-infection-induced Chlamydia trachomatis persistence is not mediated by any known persistence inducer or anti-chlamydial pathway. *Microbiology,* 154**,** 971-8.

VOGEL, J. & LUISI, B. F. 2011. Hfq and its constellation of RNA. *Nat Rev Microbiol,* 9**,** 578-89.

VORIMORE, F., HSIA, R. C., HUOT-CREASY, H., BASTIAN, S., DERUYTER, L., PASSET, A., SACHSE, K., BAVOIL, P., MYERS, G. & LAROUCAU, K. 2013. Isolation of a New Chlamydia species from the Feral Sacred Ibis (Threskiornis aethiopicus): Chlamydia ibidis. *PLoS One,* 8**,** e74823.

WAGNER, E. G. 2013. Cycling of RNAs on Hfq. *RNA Biol,* 10**,** 619-26.

WASSARMAN, K. M. 2007. 6S RNA: a regulator of transcription. *Mol Microbiol,* 65**,** 1425-31.

WATERS, L. S. & STORZ, G. 2009. Regulatory RNAs in bacteria. *Cell,* 136**,** 615-28.

WHO. 2019. *Sexually transmitted infections (STIs)* [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/sexually-transmitted-infections-(stis) [Accessed].

WICHLAN, D. G. & HATCH, T. P. 1993. Identification of an early-stage gene of Chlamydia psittaci 6BC. *J Bacteriol,* 175**,** 2936-42.

WINDBICHLER, N., VON PELCHRZIM, F., MAYER, O., CSASZAR, E. & SCHROEDER, R. 2008. Isolation of small RNA-binding proteins from E. coli: Evidence for frequent interaction of RNAs with RNA polymerase. *RNA Biology,* 5**,** 30-40.

WYRICK, P. B., CHOONG, J., DAVIS, C. H., KNIGHT, S. T., ROYAL, M. O., MASLOW, A. S. & BAGNELL, C. R. 1989. Entry of genital Chlamydia trachomatis into polarized human epithelial cells. *Infect Immun,* 57**,** 2378-89.

YAN, B., BOITANO, M., CLARK, T. A. & ETTWILLER, L. 2018. SMRT-Cappable-seq reveals complex operon variants in bacteria. *Nature Communications,* 9**,** 3676.

YU, H. H. & TAN, M. 2003. Sigma28 RNA polymerase regulates hctB, a late developmental gene in Chlamydia. *Mol Microbiol,* 50**,** 577-84.

YU, H. H. Y., KIBLER, D. & TAN, M. 2006. In silico prediction and functional validation of sigma28-regulated genes in Chlamydia and Escherichia coli. *Journal of bacteriology,* 188**,** 8206-8212.

ZHANG, J. P. & STEPHENS, R. S. 1992. Mechanism of C. trachomatis attachment to eukaryotic host cells. *Cell,* 69**,** 861-9.

ZHANG, L., DOUGLAS, A. L. & HATCH, T. P. 1998. Characterization of a Chlamydia psittaci DNA binding protein (EUO) synthesized during the early and middle phases of the developmental cycle. *Infect Immun,* 66**,** 1167-73.

ZHANG, L., HOWE, M. M. & HATCH, T. P. 2000. Characterization of in vitro DNA binding sites of the EUO protein of Chlamydia psittaci. *Infect Immun,* 68**,** 1337-49.

ZHAO, J., OHSUMI, T. K., KUNG, J. T., OGAWA, Y., GRAU, D. J., SARMA, K., SONG, J. J., KINGSTON, R. E., BOROWSKY, M. & LEE, J. T. 2010. Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Molecular Cell,* 40**,** 939-953.

7. Appendix

7.1. Abbreviations

	%	percent
	°C	celsius
	µ	micro
	α	alpha
	β	beta
	σ	sigma
	5'P	5' monophosphate
	5'PPP	5' monophosphate
A	A	adenine (nucleobase)
	AB	aberrant body
	$AgNO_3$	Silver nitrate
	APS	Ammonium persulfate
	ATP	Adenosine triphosphate
B	*B. subtilis*	*Bacillus subtilis*
	bp	base pair
	BSA	Bovine serum albumin
C	C	cytosine (nucleobase)
	C. difficile	Clostridioides difficile
	C. muridarium	Chlamydia muridarium
	C. pneumoniae	*Chlamydia pneumoniae*
	C. psittaci	*Chlamydia psittaci*
	C. trachomatis	*Chlamydia trachomatis*
	$CaCl_2$	Calcium chloride
	cDNA	complementary DNA
	CDS	Chlamydial coding transcripts
	cm^2	square centimetre
	CO_2	Carbon dioxide
	ctr	*Chlamydia trachomatis*
	Cys	Cysteine
D	DEPC	diethylpyrocarbonate
	dH_2O	Distilled water
	DMSO	Dimethyl sulfoxide
	DNA	Deoxyribonucleic acid
	dRNA-Seq	Differential RNA-sequencing
	DTT	Dithiothreito
E	*E. coli*	*Escherichia coli*
	e.g.	latin *exempli gratia*
	EB	Elementary body
	ECL	Enhanced chemiluminescence
	EDTA	ethylene-diamine-tetraacetic acid
	et al.	latin *et alia*
	EtBr	Ethidium bromide
	EtOH	Ethanol

F	FCS	Fetal calf serum
G	G	guanine (nucleobase)
	g	gram
	Gln	Glutamine
H	hCDS	Human transcripts
	HCl	Hydrogen chloride
	hpi	hours post infection
	hr	hour
I	IFU	inclusion forming unit
	IS	insertion
K	kb	kilo base pairs
	KCl	Potassium chloride
	kDA	Kilo Dalton
	KH_2PO_4	Monopotassium phosphate
L	l	liter
	LB	lysogeny broth
	LGV	lympho granuloma venerum
	log	Logarithm
M	M	molar
	m/s	meter per second
	mA/cm^2	milliampere per square centimetre
	Max	maximum
	Mb	mega base pairs
	Met	Methionine
	$MgCl_2$	Magnesium chloride
	$MgSO_4$	Magnesium sulfate
	min	minute (s)
	min	minimum
	ml	millilitre
	mm	millimeter
	MOI	multiplicity of infection
	mRNA	messenger RNA
	MS	Mass spectrometry
N	Na_2CO_3	Sodium carbonate
	Na_2HPO_4	Disodium phosphate
	$Na_2S_2O_3$	Sodium thiosulfate
	NaCl	Sodium chloride
	NaOAc	Sodium acetate
	ncRNA	non-coding RNA
	ng	nanogram
	nm	nanometer
	No.	number
	nt	nucleotide
O	OD600	Optical density measured at a wavelength of 600 nm
P	P/C/I	phenol/chloroform/isoamyl alcohol
	PAGE	polyacrylamide gel electrophoresis

	PBS	phosphate buffered saline
	PCA	principle component analysis
	PCR	Polymerase chain reaction
	pH	lat. *potentia hydrogeneii*
	PMSF	Phenylmethylsulfonylfluorid
	PSS	processing site
	PVDF	Polyvinylidene difluoride
R	RB	reticular bodies
	RBP	RNA binding protein
	RNA	ribonucleic acid
	RNAP	RNA polymerase
	rpm	revolutions per minute
	rRNA	ribosomal RNA
S	s	second (s)
	SDS	sodium dodecyl sulfate
	SOC	Super optimal broth with catabolite repression
	SPG	sucrose-phosphate-glutamate buffer
	sRNA	small RNA
	SRP	signal recognition particle
T	T	Thymine (nucleobase)
	TBS	Tris buffered saline
	TBS-T	Tris buffered saline with Tween-20
	TEMED	Tetramethylethylenediamine
	tmRNA	transfer messenger RNA
	tRNA	transfer RNA
	TSS	transcriptional start sites
U	U	Uracil (nucleobase)
	UPSS	processing site with unknown gene
	UTR	untranslated region
	UTRP	untranslated region from Processing Event
	UTSS	transcriptional start with unknown gene
	UV-VIS	Ultraviolet–visible
V	V	volt
	v/v	volume per volume
W	w/v	weight per volume

7.2. List of figures

7.3. List of Tables

7.4. Supplementary information

7.4.1. TagRNA-seq trimming

Following command was used for trimming of the TagRNA-seq raw data:

```
trimmomatic-0.36.jar SE
-phred33
ILLUMINACLIP:adapter:2:30:10
SLIDINGWINDOW:4:15
LEADING:3
TRAILING:3
MINLEN:30
```

The adapters sequences used for trimming of the TagRNA-seq raw data:

```
>TruSeq_3'fwd
AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
>TruSeq-3'rev
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
>TruSeq_5'fwd
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA
>TruSeq5'rev
ACACTCTTTCCCTACACGACGCTCTTCCGATCT
>PolyA80
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAA
>PolyT80
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TTT
```

7.4.2. TagRNA-seq processing script

The following script was written for processing of the TagRNA-seq Data and annotation generation:

```
from Bio import SeqIO
import numpy as np
import pandas as pd
import gffutils
import re
import matplotlib.pyplot as plt
get_ipython().run_line_magic('matplotlib', 'inline')
# Daten laden Ã¼ber funktion und in ein pandas Array ziehen
# Annotation erstellen und laden in db
```

```
#gffutils.create_db("AM884176.1.gff",'chlamydiadatabase.db', force=True,
keep_order=True, merge_strategy='merge', sort_attribute_values=True)
db = gffutils.FeatureDB('chlamydiadatabase.db', keep_order=True)
# Dateien laden laden "tracks" TSS PSS und unassigned je plus und minus
# Sequence laden von fasta file
def get_fasta_sequence (fastafile):
   fasta_sequences = list(SeqIO.parse(open(fastafile),'fasta'))
   allletters=""
   for fasta in fasta_sequences:
      allletters = allletters  +fasta.seq
   return (pd.Series(list(allletters)))
#get_fasta_sequence("434bu_chr+pl2.fa")
# Dateien laden laden "tracks" TSS PSS und unassigned je plus und minus
# GroÃŸe Tabelle erstellen mit allen Rohdaten

def combine_Trackdata
(fasta,tSSplusFile,pSSplusFile,tSSminusFile,pSSminusFile,unassignedPLUSFile,unassignedMin
usFile):
   completteDataFrame=pd.concat([fasta,
                     tSSplusFile['#RefName'],
                     tSSplusFile['Pos'],
                     tSSplusFile['Coverage'],
                     pSSplusFile['Coverage'],
                     tSSminusFile['Coverage'],
                     pSSminusFile['Coverage'],
                     unassignedPLUSFile['Coverage'],
                     unassignedMinusFile['Coverage']],axis=1)
   completteDataFrame.columns=['letter','RefName','Pos','TSS+','PSS+','TSS-','PSS-
','Unassigned+','Unassigned-']
   return completteDataFrame

def load_track (trackfile):
   return pd.read_csv(trackfile ,header=0 ,sep="\t")

def filter_tracktable_by_Referenz (completteDataFrame,Referenz):
   return completteDataFrame.loc[completteDataFrame['RefName'] == Referenz]
# GroÃŸe Tabelle filtern mit Threshold(50)
def filter_track_data(trackdatatable,threshold):
   for tracks in list(trackdatatable.columns.values):
      match= re.match("[TP]SS[+-]",tracks)
      if match:
         trackdatatable.loc[trackdatatable[tracks] <= threshold,[tracks]]=1


# Prozentualle Ã¤nderung messen,anhÃ¤ngen & filtern per threshold (0.7) beachten fÃ¼r
MINUS-TRACK (-0.7)
# Bei jedem Datenpunkt umfeld(Threshold +/- 10 bp) beachten und nur Laut Rohdaten,den
Messpunkt mit dem maxima Ã¼bernehmen
```

```
def calculate_procent_change_table(trackdatatable):
   return(pd.concat([trackdatatable['TSS+'].pct_change(),
                trackdatatable['PSS+'].pct_change(),
                trackdatatable['TSS-'].pct_change(),
                trackdatatable['PSS-'].pct_change()],axis=1))

def filter_procent_TSS_data(procentChangeTablle,threshold):
   threshold=abs(threshold)
   for tracks in list(procentChangeTablle.columns.values):
      match= re.match("TSS[+-]",tracks)
      if match:
         plusOrMinus=re.match("TSS[+]",tracks)
         if plusOrMinus:
            procentChangeTablle.loc[procentChangeTablle[tracks] <= threshold,[tracks]]=0
         else:
            procentChangeTablle.loc[procentChangeTablle[tracks] >= threshold*-1,[tracks]]=0

def check_surrounding_TSS_procent(procentChangeTablle,tracktable,basesAround):
   for tracks in list(procentChangeTablle.columns.values):
      match= re.match("TSS[+-]",tracks)
      if match:
         indexes = procentChangeTablle.index[procentChangeTablle[tracks] != 0].tolist()
         count=1
         while count < len(indexes):
            if indexes[count] in range(indexes[count-1]-basesAround,indexes[count-
1]+basesAround,1):
               if tracktable[tracks].iloc[indexes[count-1]] <
tracktable[tracks].iloc[indexes[count]]:
                  procentChangeTablle[tracks].iloc[indexes[count-1]]=0
               else:
                  procentChangeTablle[tracks].iloc[indexes[count]]=0
            count+=1

# Analyse der PSS wie TSS
def filter_procent_PSS_data(procentChangeTablle,threshold):
   threshold=abs(threshold)
   for tracks in list(procentChangeTablle.columns.values):
      match= re.match("PSS[+-]",tracks)
      if match:
         plusOrMinus=re.match("PSS[+]",tracks)
         if plusOrMinus:
            procentChangeTablle.loc[procentChangeTablle[tracks] <= threshold,[tracks]]=0
         else:
            procentChangeTablle.loc[procentChangeTablle[tracks] >= threshold*-1,[tracks]]=0

def check_surrounding_PSS_procent(procentChangeTablle,tracktable,basesAround):
   for tracks in list(procentChangeTablle.columns.values):
      match= re.match("PSS[+-]",tracks)
```

```
      if match:
         indexes = procentChangeTablle.index[procentChangeTablle[tracks] != 0].tolist()
         count=1
         while count < len(indexes):
            if indexes[count] in range(indexes[count-1]-basesAround,indexes[count-
1]+basesAround,1):
               if tracktable[tracks].iloc[indexes[count-1]] <
tracktable[tracks].iloc[indexes[count]]:
                  procentChangeTablle[tracks].iloc[indexes[count-1]]=0
               else:
                  procentChangeTablle[tracks].iloc[indexes[count]]=0
            count+=1


# AbhÃ¤nig von umfeld (threshold +/- 10 bp) von den Rohwerten PSS oder TSS ausgeben
def identify_TSS_or_PSS(table,raw,basesAround):
   TSSindexes = table.index[table["TSS+"] != 0].tolist()
   PSSindexes = table.index[table["PSS+"] != 0].tolist()
   count=0
   howmany=0
   while count < len(TSSindexes):
      innercount=0
      TSSrange=range(TSSindexes[count]-basesAround,TSSindexes[count]+basesAround+1,1)
      while innercount < len(PSSindexes):
         if PSSindexes[innercount] in TSSrange:
            if raw["TSS+"].iloc[TSSindexes[count]] > raw["PSS+"].iloc[PSSindexes[innercount]]:
               table["PSS+"].iloc[PSSindexes[innercount]]=0
            else:
               table["TSS+"].iloc[PSSindexes[innercount]]=0
            howmany+=1
         innercount+=1
      howmany=0
      count+=1

   TSSindexes = table.index[table["TSS-"] != 0].tolist()
   PSSindexes = table.index[table["PSS-"] != 0].tolist()
   count=0
   howmany=0
   while count < len(TSSindexes):
      innercount=0
      TSSrange=range(TSSindexes[count]-basesAround,TSSindexes[count]+basesAround+1,1)
      while innercount < len(PSSindexes):
         if PSSindexes[innercount] in TSSrange:
            if raw["TSS-"].iloc[TSSindexes[count]] > raw["PSS-"].iloc[PSSindexes[innercount]]:
               table["PSS-"].iloc[PSSindexes[innercount]]=0
            else:
               table["TSS-"].iloc[PSSindexes[innercount]]=0
            howmany+=1
```

```
            innercount+=1

        howmany=0
        count+=1

def write_table(procent,outputfile,associatedGeneDistance):
    with open(outputfile, 'w') as the_file:
        the_file.write("Position\ttrack\tDistance to gene
start\tcontig\tsource\ttype\tstart\tend\tscore\tstrand\tframe\tattributes\n")
        for track in ['TSS','PSS']:
            for direction in ['+','-']:
                specfictrack=track+direction
                indexes = procent.index[procent[specfictrack] != 0].tolist()
                count=0
                if direction=='+':
                    while count < len(indexes):
                        wasinloop=0
                        for i in db.region(seqid="AM884176.1",start=(indexes[count]-
associatedGeneDistance),end=(indexes[count]+associatedGeneDistance),strand=direction,fe
aturetype="gene"):
                            the_file.write(str(indexes[count])+"\t"+specfictrack +"\t"
+str(indexes[count]-i.start)+"\t" +str(i)+"\n")
                            wasinloop=1
                        if wasinloop==0:
                            the_file.write(str(indexes[count])+"\t"+specfictrack+"\n")
                        count+=1
                else:
                    while count < len(indexes):
                        wasinloop=0
                        for i in db.region(seqid="AM884176.1",start=(indexes[count]-
associatedGeneDistance),end=(indexes[count]+associatedGeneDistance),strand=direction,fe
aturetype="gene"):
                            the_file.write(str(indexes[count])+"\t"+specfictrack +"\t" +str(i.end-
indexes[count])+"\t" +str(i)+"\n")
                            wasinloop=1
                        if wasinloop==0:
                            the_file.write(str(indexes[count])+"\t"+specfictrack+"\n")
                        count+=1

def write_gff_rows(procent,outputfile,genome,selector):
    if selector == 'TSS':
        tracklist=['TSS']
    elif selector == 'PSS':
        tracklist=['PSS']
    else:
        tracklist=['TSS','PSS']
    with open(outputfile, 'w') as the_file:
        for track in tracklist:
```

```
        for direction in ['+','-']:
            specfictrack=track+direction
            indexes = procent.index[procent[specfictrack] != 0].tolist()
            if direction=='+':
                count=0
            while count < len(indexes):
                the_file.write(str(genome)+"\tEMBL\tregion\t"+
                        str(indexes[count])+"\t"
                        +str(indexes[count])+"\t.\t"
                        +str(direction)+
                        "\t.\tID="+str(specfictrack)+str(count)
                        +";gene=gene"+str(count)+"_"
                        +str(track)+";Name="+str(count)+"_"+
                        str(track)+
                        ";gbkey=Gene;gene_biotype="+str(track)+"\n")
                count+=1
            else:
                count=0
                while count < len(indexes):
                    the_file.write(str(genome)+"\tEMBL\tregion\t"+
                        str(indexes[count])+"\t"
                        +str(indexes[count])+"\t.\t"
                        +str(direction)+
                        "\t.\tID="+str(specfictrack)+str(count)
                        +";gene=gene"+str(count)+"_"
                        +str(track)+";Name="+str(count)+"_"+
                        str(track)+
                        ";gbkey=Gene;gene_biotype="+str(track)+"\n")
                    count+=1

rawdata = combine_Trackdata(get_fasta_sequence("434bu_chr+pl2.fa"),
                    load_track("Taq-seq_ctr_TSS.trimmed.fastqperBase1.graph"),
                    load_track("Taq-seq_ctr_PSS.trimmed.fastqperBase1.graph"),
                    load_track("Taq-seq_ctr_TSS.trimmed.fastqperBase2.graph"),
                    load_track("Taq-seq_ctr_PSS.trimmed.fastqperBase2.graph"),
                    load_track("Taq-seq_ctr_unassigned.trimmed.fastqperBase1.graph"),
                    load_track("Taq-seq_ctr_unassigned.trimmed.fastqperBase2.graph")
                    )
rawdata=filter_tracktable_by_Referenz(rawdata,"AM884176.1")
filter_track_data(rawdata,50)
###TSS
procentchange=calculate_procent_change_table(rawdata)
filter_procent_TSS_data(procentchange,0.7)
check_surrounding_TSS_procent(procentchange,rawdata,20)
###PSS
filter_procent_PSS_data(procentchange,0.7)
check_surrounding_PSS_procent(procentchange,rawdata,20)
identify_TSS_or_PSS(procentchange,rawdata,30)
```

```
print("FINAL TSS and PSS")
print("TSS on plus strand :" + str(len(procentchange.index[procentchange["TSS+"] !=
0].tolist())))
print("TSS on minus strand :" + str(len(procentchange.index[procentchange["TSS-"] !=
0].tolist())))
print("PSS on plus strand :" + str(len(procentchange.index[procentchange["PSS+"] !=
0].tolist())))
print("PSS on minus strand :" + str(len(procentchange.index[procentchange["PSS-"] !=
0].tolist())))

write_table(procentchange,"outputfile.tsv",100)
write_gff_rows(procentchange,"test.gff","AM884176.1","both")
write_gff_rows(procentchange,"test_TSS.gff","AM884176.1","TSS")
write_gff_rows(procentchange,"test_PSS.gff","AM884176.1","PSS")

def make_annoation_for_TSS_with_gene(procent,outfile,UTRzone):
   tracklist=['TSS']
   with open(outfile, 'w') as the_file:
      for track in tracklist:
         for direction in ['+','-']:
            specfictrack=track+direction
            indexes = procent.index[procent[specfictrack] != 0].tolist()
            indexes_used_for_genes=[]
            if direction=='+':
               for i in db.region(seqid="AM884176.1",strand=direction,featuretype="gene"):
                     indexes_within_genes=count(indexes,i.start-UTRzone,i.end)
                     indexes_used_for_genes.extend(indexes_within_genes)
                     for e in indexes_within_genes:
                        if (e<i.start):

the_file.write("AM884176.1"+"\ttaqseq\tUTR\t"+str(e)+"\t"+str(i.start)+"\t.\t+\t.\tID=TSSp"
+str(i.attributes['Name'][0])+"_"+str(e)+";Note=TSSp"+str(i.attributes['Name'][0])+"_"+str(e)
+";gene_biotype=UTR+\n")
                        else:

the_file.write("AM884176.1"+"\ttaqseq\tTSS\t"+str(e)+"\t"+str(i.end)+"\t.\t+\t.\tID=TSSp"+s
tr(i.attributes['Name'][0])+"_"+str(e)+";Note=TSSp"+str(i.attributes['Name'][0])+"_"+str(e)+";
gene_biotype=TSS+\n")
               for e in list(set(indexes)-set(indexes_used_for_genes)):

the_file.write("AM884176.1"+"\ttaqseq\tUTSS\t"+str(e)+"\t"+str(e+UTRzone)+"\t.\t+\t.\tID=
TSSp"+str(e)+";Note=TSSp"+str(e)+";gene_biotype=UTSS+\n")
            else:
               for i in db.region(seqid="AM884176.1",strand=direction,featuretype="gene"):
                     indexes_within_genes=count(indexes,i.start,i.end+UTRzone)
                     indexes_used_for_genes.extend(indexes_within_genes)
                     for e in indexes_within_genes:
```

```
                    if (e>i.end):

the_file.write("AM884176.1"+"\ttaqseq\tUTR\t"+str(i.end)+"\t"+str(e)+"\t.\t-
\t.\tID=TSSm"+str(i.attributes['Name'][0])+"_"+str(e)+";Note=TSSm"+str(i.attributes['Name']
[0])+"_"+str(e)+";gene_biotype=UTR-\n")
                    else:

the_file.write("AM884176.1"+"\ttaqseq\tTSS\t"+str(i.start)+"\t"+str(e)+"\t.\t-
\t.\tID=TSSm"+str(i.attributes['Name'][0])+"_"+str(e)+";Note=TSSm"+str(i.attributes['Name']
[0])+"_"+str(e)+";gene_biotype=TSS-\n")
            for e in list(set(indexes)-set(indexes_used_for_genes)):
                the_file.write("AM884176.1"+"\ttaqseq\tUTSS\t"+str(e-
UTRzone)+"\t"+str(e)+"\t.\t-
\t.\tID=TSSm"+str(e)+";Note=TSSm"+str(e)+";gene_biotype=UTSS-\n")
        print("All postions for ",track," ",direction ,len(indexes))
        print("Used postions for ",track," ",direction ,len(indexes_used_for_genes))
        print("Amount of unsed for ",track," ",direction ,len(list(set(indexes)-
set(indexes_used_for_genes))))

def make_annoation_for_PSS_with_gene(procent,outfile,extend):
  tracklist=['PSS']
  with open(outfile, 'w') as the_file:
    for track in tracklist:
      for direction in ['+','-']:
        specfictrack=track+direction
        indexes = procent.index[procent[specfictrack] != 0].tolist()
        indexes_used_for_genes=[]
        if direction=='+':
          for i in db.region(seqid="AM884176.1",strand=direction,featuretype="gene"):
              indexes_within_genes=count(indexes,i.start,i.end)
              indexes_used_for_genes.extend(indexes_within_genes)
              if indexes_within_genes:
                gene_array=[i.start]
                gene_array.extend(indexes_within_genes)
                gene_array.extend([i.end])
                gene_array.sort()
                counting=0
                while counting < (len(gene_array)-1):

the_file.write("AM884176.1"+"\ttaqseq\tPSS\t"+str(gene_array[counting])+"\t"+str(gene_ar
ray[counting+1])+"\t.\t+\t.\tID=PSS+_"+str(i.attributes['Name'][0])+"_"+str(counting)+";Note
=PSS+_"+str(i.attributes['Name'][0])+"_"+str(counting)+";gene_biotype=PSS+\n")
                    counting+=1
            for e in list(set(indexes)-set(indexes_used_for_genes)):
                the_file.write("AM884176.1"+"\ttaqseq\tUPSS\t"+str(e-
extend)+"\t"+str(e+extend)+"\t.\t+\t.\tID=PSSp"+str(e)+";Note=PSSp"+str(e)+";gene_biotype
=UPSS+\n")
        else:
```

```
            for i in db.region(seqid="AM884176.1",strand=direction,featuretype="gene"):
                indexes_within_genes=count(indexes,i.start,i.end)
                indexes_used_for_genes.extend(indexes_within_genes)
                if indexes_within_genes:
                   gene_array=[i.start]
                   gene_array.extend(indexes_within_genes)
                   gene_array.extend([i.end])
                   gene_array.sort()
                   counting=0
                   while counting < (len(gene_array)-1):

the_file.write("AM884176.1"+"\ttaqseq\tPSS\t"+str(gene_array[counting])+"\t"+str(gene_ar
ray[counting+1])+"\t.\t-\t.\tID=PSS-
_"+str(i.attributes['Name'][0])+"_"+str(counting)+";Note=PSS-
_"+str(i.attributes['Name'][0])+"_"+str(counting)+";gene_biotype=PSS-\n")
                         counting+=1
            for e in list(set(indexes)-set(indexes_used_for_genes)):
               the_file.write("AM884176.1"+"\ttaqseq\tUPSS\t"+str(e-
extend)+"\t"+str(e+extend)+"\t.\t-
\t.\tID=PSSm"+str(e)+";Note=PSSm"+str(e)+";gene_biotype=UPSS-\n")
         print("All postions for ",track," ",direction ,len(indexes))
         print("Used postions for ",track," ",direction ,len(indexes_used_for_genes))
         print("Amount of unsed for ",track," ",direction ,len(list(set(indexes)-
set(indexes_used_for_genes))))


def count(list1, l, r):
   return list(x for x in list1 if l <= x <= r)

make_annoation_for_TSS_with_gene(procentchange,"TSS.gff",100)
make_annoation_for_PSS_with_gene(procentchange,"PSS.gff",20)

rawdata = combine_Trackdata(get_fasta_sequence("434bu_chr+pl2.fa"),
                  load_track("Taq-seq_ctr_TSS.trimmed.fastqperBase1.graph"),
                  load_track("Taq-seq_ctr_PSS.trimmed.fastqperBase1.graph"),
                  load_track("Taq-seq_ctr_TSS.trimmed.fastqperBase2.graph"),
                  load_track("Taq-seq_ctr_PSS.trimmed.fastqperBase2.graph"),
                  load_track("Taq-seq_ctr_unassigned.trimmed.fastqperBase1.graph"),
                  load_track("Taq-seq_ctr_unassigned.trimmed.fastqperBase2.graph")
                  )
rawdata=filter_tracktable_by_Referenz(rawdata,"AM884176.1")
filter_track_data(rawdata,50)
###TSS
procentchange=calculate_procent_change_table(rawdata)
filter_procent_TSS_data(procentchange,0.7)
check_surrounding_TSS_procent(procentchange,rawdata,50)
###PSS
filter_procent_PSS_data(procentchange,0.7)
```

```
check_surrounding_PSS_procent(procentchange,rawdata,50)
identify_TSS_or_PSS(procentchange,rawdata,30)

print("FINAL TSS and PSS")
print("TSS on plus strand :" + str(len(procentchange.index[procentchange["TSS+"] !=
0].tolist())))
print("TSS on minus strand :" + str(len(procentchange.index[procentchange["TSS-"] !=
0].tolist())))
print("PSS on plus strand :" + str(len(procentchange.index[procentchange["PSS+"] !=
0].tolist())))
print("PSS on minus strand :" + str(len(procentchange.index[procentchange["PSS-"] !=
0].tolist())))

write_table(procentchange,"outputfile_broader.tsv",100)
write_gff_rows(procentchange,"test_broader.gff","AM884176.1","both")
write_gff_rows(procentchange,"test_broader_TSS.gff","AM884176.1","TSS")
write_gff_rows(procentchange,"test_broader_PSS.gff","AM884176.1","PSS")
```

7.4.3. Additional Results of the Motif search of the TagRNA-seq data

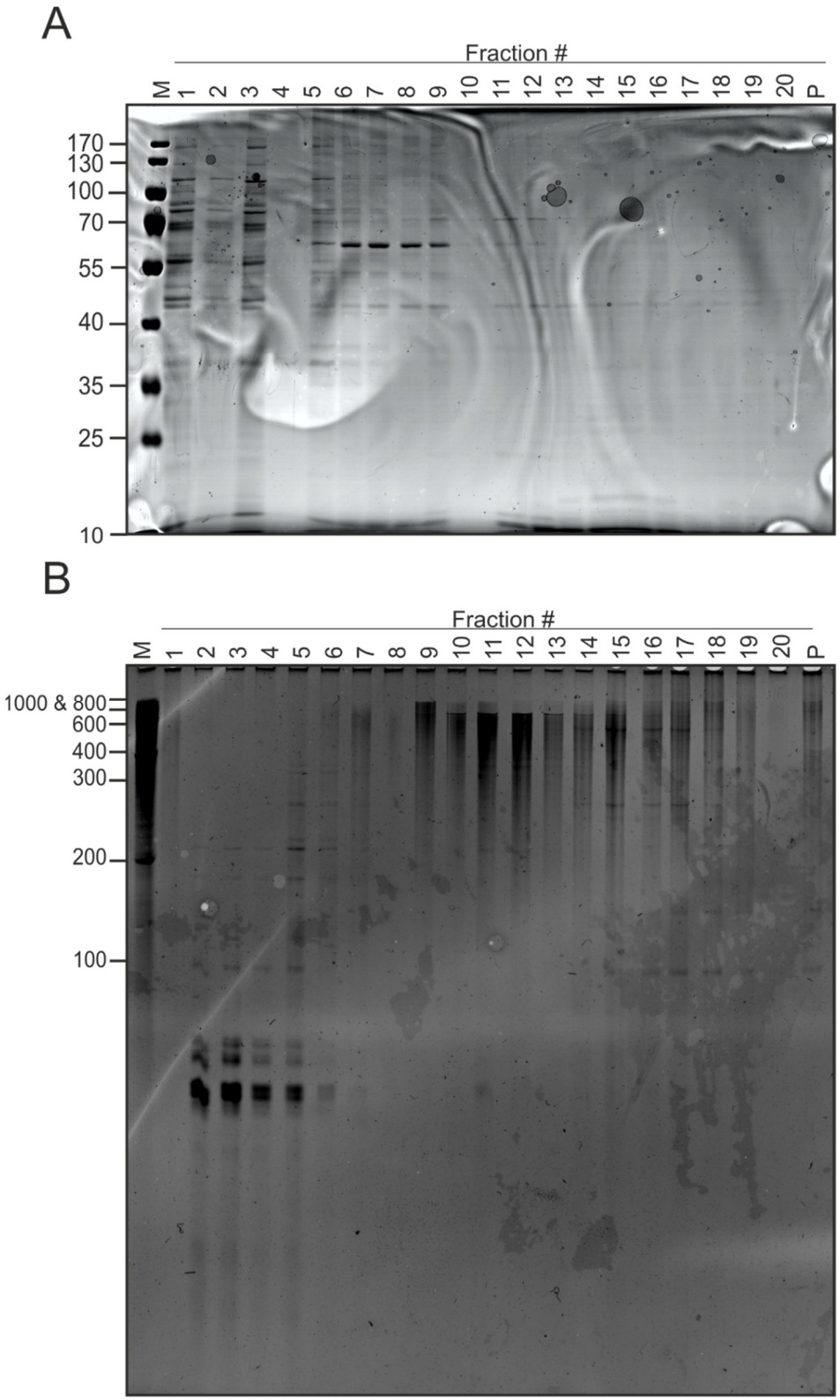
Appendix table 23: Results from tomtom from the MEME tool suit searched against all prokaryotic DNA databases for the third motif found using the 40 bp upstream sequences from the new annotated intergenic regions.

Motif (Database entry)	p-value	E-value	q-value
Fur *V.cholerae* (EXPREG_000008b0)	3.78E-05	1.86E-02	3.66E-02
Fur *P.aeruginosa*(EXPREG_00000c80)	1.24E-04	6.13E-02	5.65E-02
Fur (MX000013)	1.77E-04	8.73E-02	5.65E-02
Fur (MX000028)	2.33E-04	1.15E-01	5.65E-02
PerR *B.subtilis* (EXPREG_00000bc0)	7.28E-04	3.59E-01	1.12E-01
rpoD (rpoD18)	1.45E-03	7.16E-01	1.41E-01
Fur *B.subtilis* (EXPREG_00000b40)	1.51E-03	7.47E-01	1.41E-01
Fnr (MX000004)	1.60E-03	7.92E-01	1.41E-01
Fur *A.salmonicida* (EXPREG_00000360)	1.92E-03	9.47E-01	1.45E-01
Fur *Y.pestis* (EXPREG_00000a20)	1.94E-03	9.58E-01	1.45E-01
fur (fur)	3.11E-03	1.53E+00	2.01E-01
Fur *S.enterica* (EXPREG_00001020)	3.11E-03	1.54E+00	2.01E-01
Fnr (Gammaproteobacteria)	4.08E-03	2.01E+00	2.47E-01
NikR *H.pylori* (EXPREG_00000670)	4.82E-03	2.38E+00	2.64E-01
Fur *L.monocytogenes* (EXPREG_000013d0)	4.91E-03	2.43E+00	2.64E-01
Fur *N.gonorrhoeae* (EXPREG_00000ec0)	5.76E-03	2.84E+00	2.64E-01
ArcA (MX000091)	6.31E-03	3.12E+00	2.64E-01
Fur *P.syringae* (EXPREG_00001110)	6.54E-03	3.23E+00	2.64E-01
SigE (MX000068)	8.55E-03	4.22E+00	3.28E-01
argR (argR)	1.09E-02	5.38E+00	3.64E-01
AlgU (MX000035)	1.40E-02	6.89E+00	4.48E-01
SigB (MX000072)	1.65E-02	8.13E+00	4.99E-01
MalR C.perfringens (MalR_Clostridium_perfringens)	1.76E-02	8.67E+00	5.16E-01
Fur A.ferrooxidans (EXPREG_00000370)	1.84E-02	9.08E+00	5.24E-01

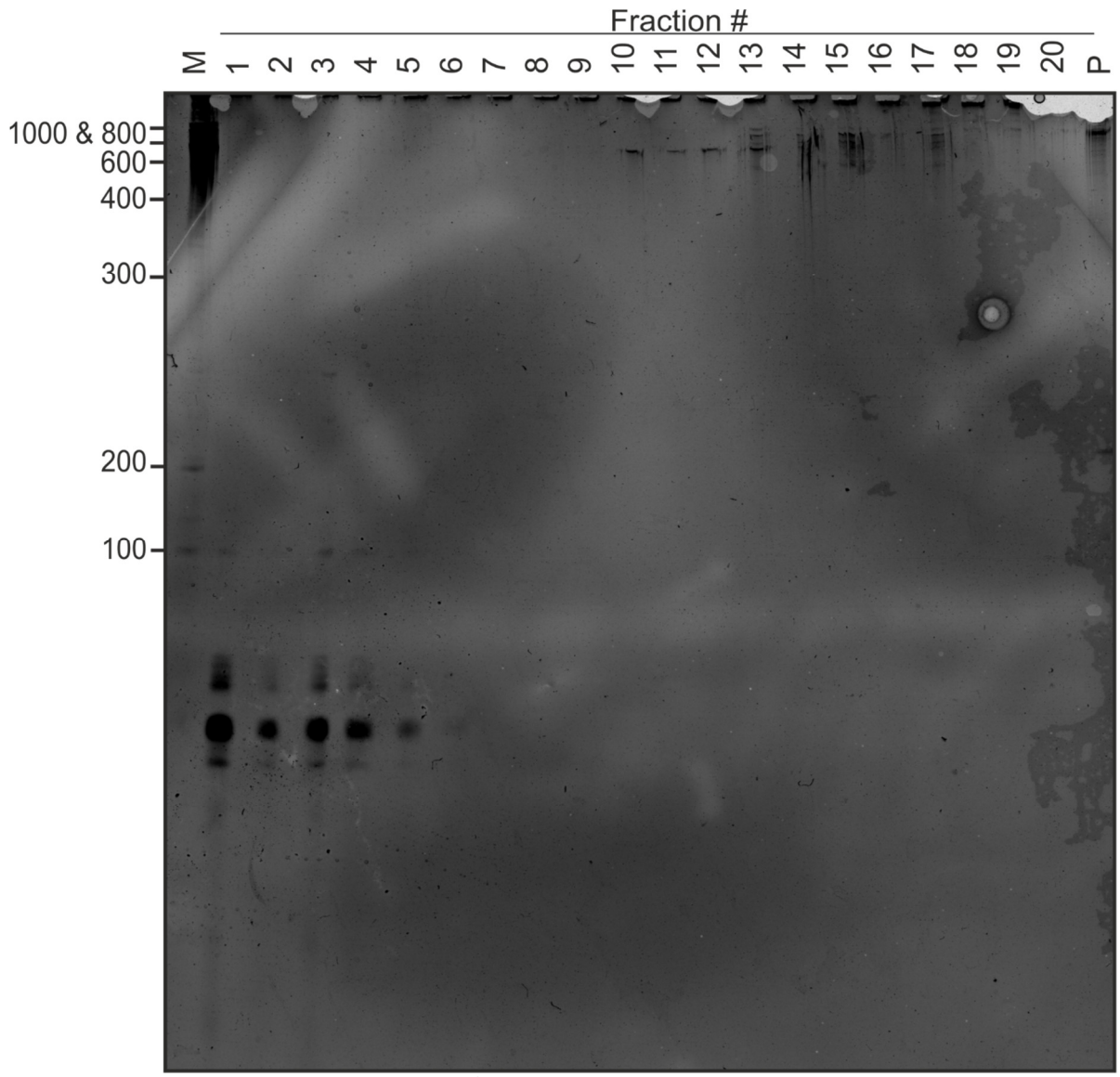
Appendix table 24: Results from tomtom from the MEME tool suit searched against all prokaryotic DNA databases for the fourth motif found using the 40 bp upstream sequences from the new annotated intergenic regions.

Motif (Database entry)	p-value	E-value	q-value
LexA *M.tuberculosis* (EXPREG_000005d0)	3.87E-03	1.91E+00	1.00E+00
LexA *C.glutamicum* (EXPREG_00000020)	4.91E-03	2.43E+00	1.00E+00
DinR/LexA (MX000025)	5.18E-03	2.56E+00	1.00E+00
Atu5393_Rhizobiales	1.15E-02	5.70E+00	1.00E+00
RSc0472_Burkholderiales	1.18E-02	5.85E+00	1.00E+00
IscR_Rhodospirillales_Sphingomonadales	1.40E-02	6.91E+00	1.00E+00
FruR_Alpha_Beta_Xanthomonadales	1.41E-02	6.95E+00	1.00E+00
LexA *S.aureus* (EXPREG_00001790)	1.61E-02	7.96E+00	1.00E+00
AmrZ *P.aeruginosa* (EXPREG_000004d0)	1.69E-02	8.34E+00	1.00E+00

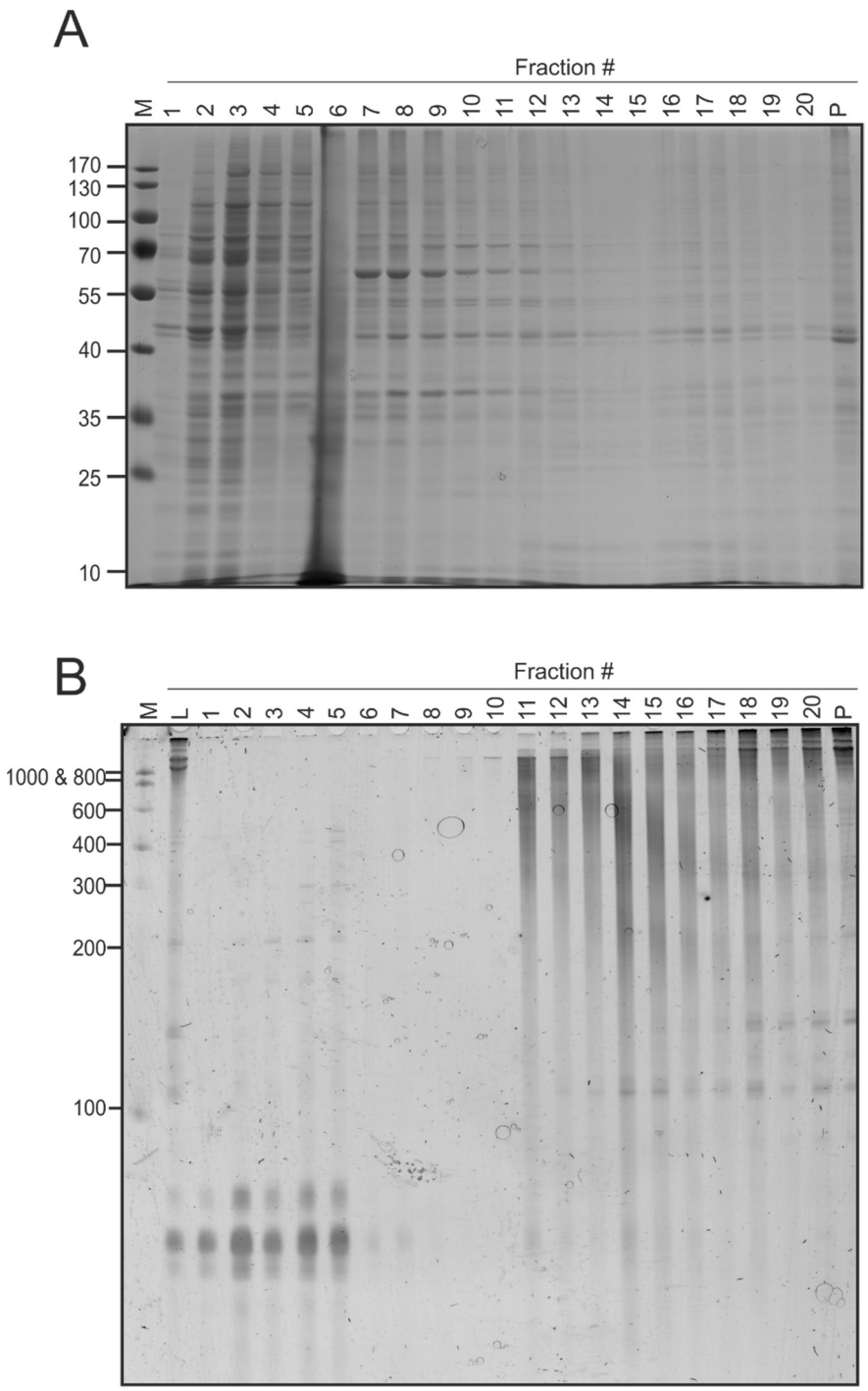
7.4.4. Additional Results of the Gradient establishment



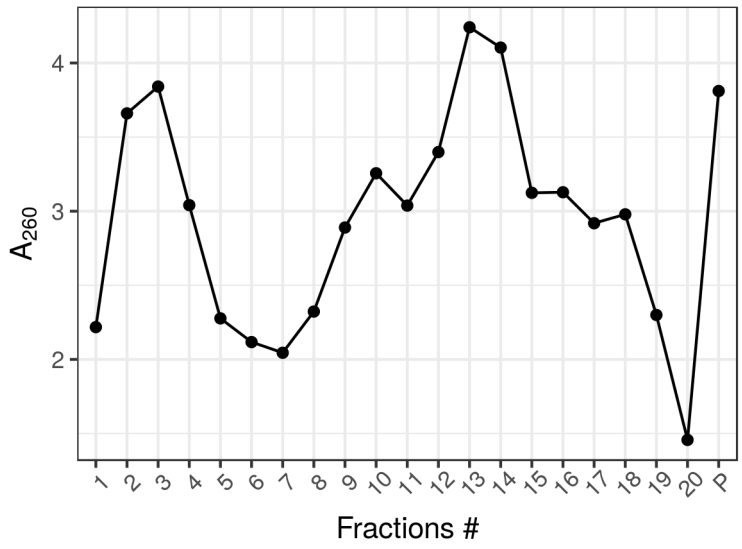
Appendix figure 37: Pellets from thirty 150 cm^2-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep at 4.0 m/s for 20 s once. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. (A) Proteins were separated using a 12 % SDS-PAGE and visualised via colloidal Coomassie staining. The first lane shows the protein marker (M) followed by the protein samples from the different fractions of the gradient collected from top (1) to pellet (P). (B) RNAs were loaded onto 7M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows RiboRuler Low range, followed by the samples of the different fractions of the gradient collected from top (1) to pellet (P).



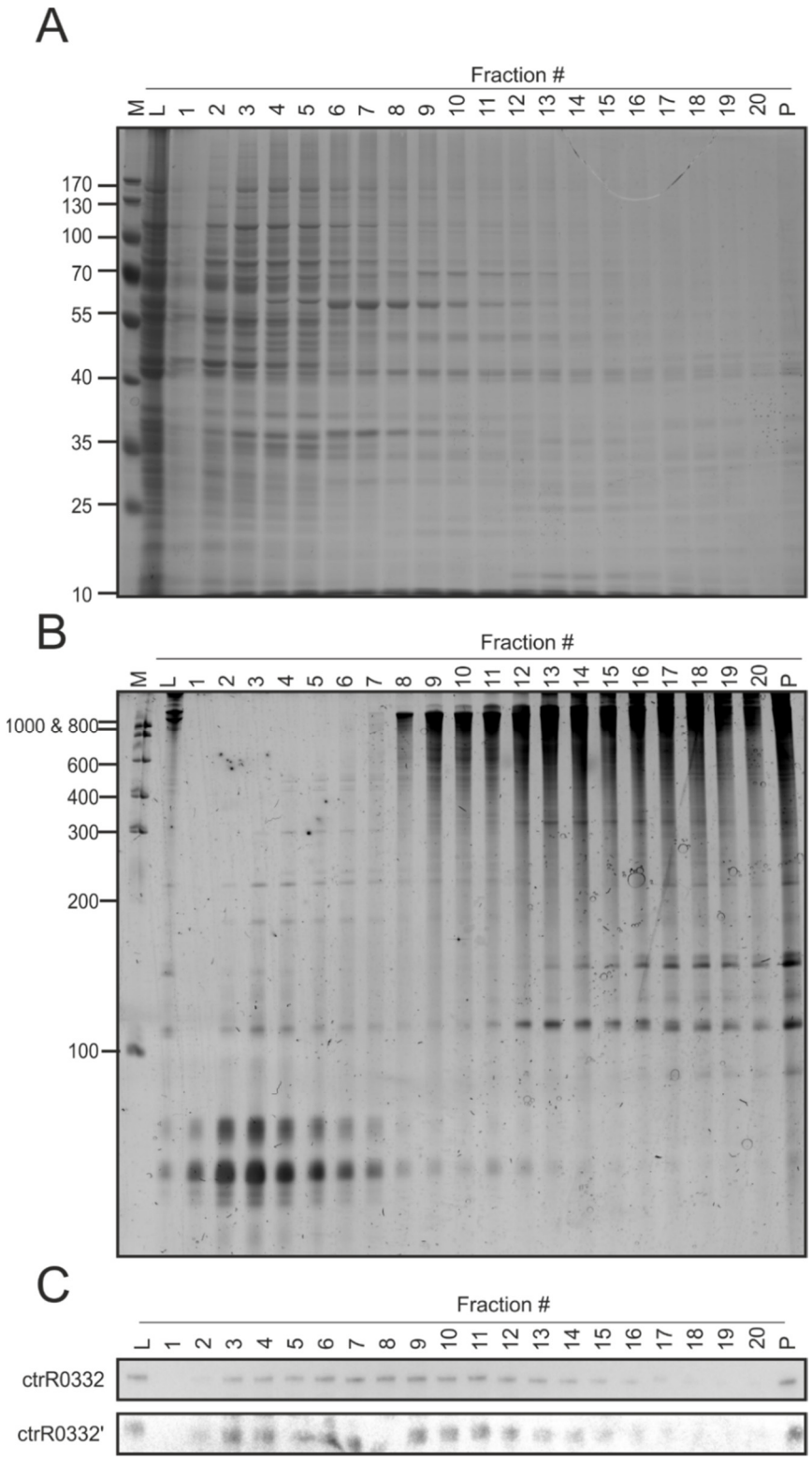
Appendix figure 38: Pellets from forty 150 cm^2-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep at 4.5 m/s for 20 s once. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. RNAs were loaded onto 7M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows RiboRuler Low range, followed by the samples of the different fractions of the gradient collected from top (1) to pellet (P).



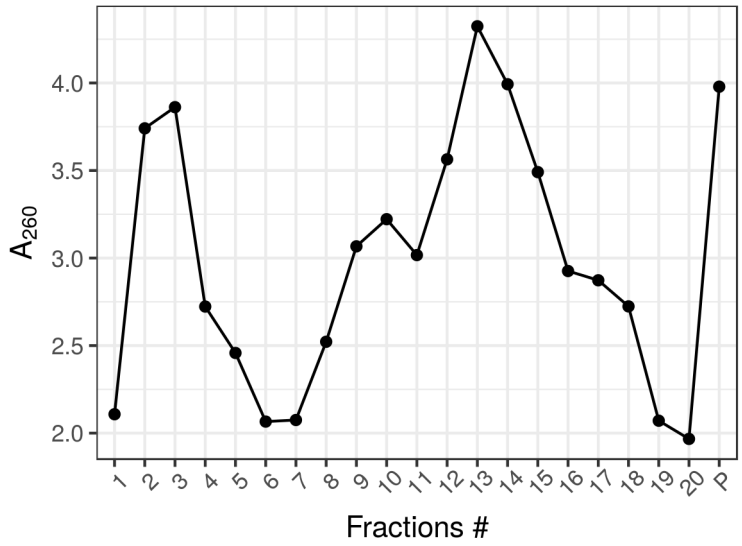
Appendix figure 39: Pellets from forty 150 cm²-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep at 4.0 m/s for 20 s once. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. (A) Proteins were separated using a 12 % SDS-PAGE and visualised via colloidal Coomassie staining. The first lane shows the protein marker (M) followed by the protein samples from the different fractions of the gradient collected from top (1) to pellet (P). (B) RNAs were loaded onto 7M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows RiboRuler Low range, followed by the samples of the different fractions of the gradient collected from top (1) to pellet (P).



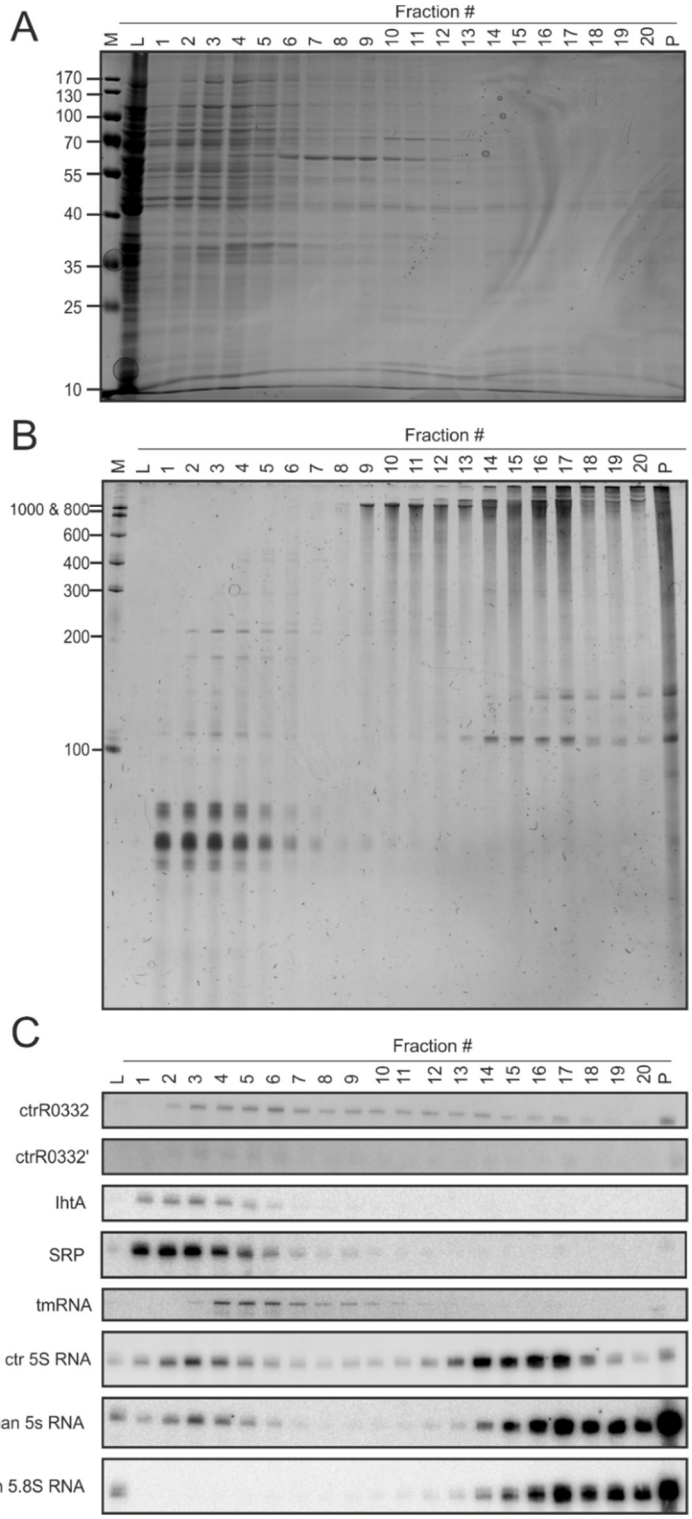
Appendix figure 40: Absorbance of all collected fractions (1-P) measured at 260nm via NanoDrop spectrophotometer. Pellets of forty 150 cm^2 dishes of cells infected with *Chlamydia* were used and lysed at 4 m/s for 20 s once.



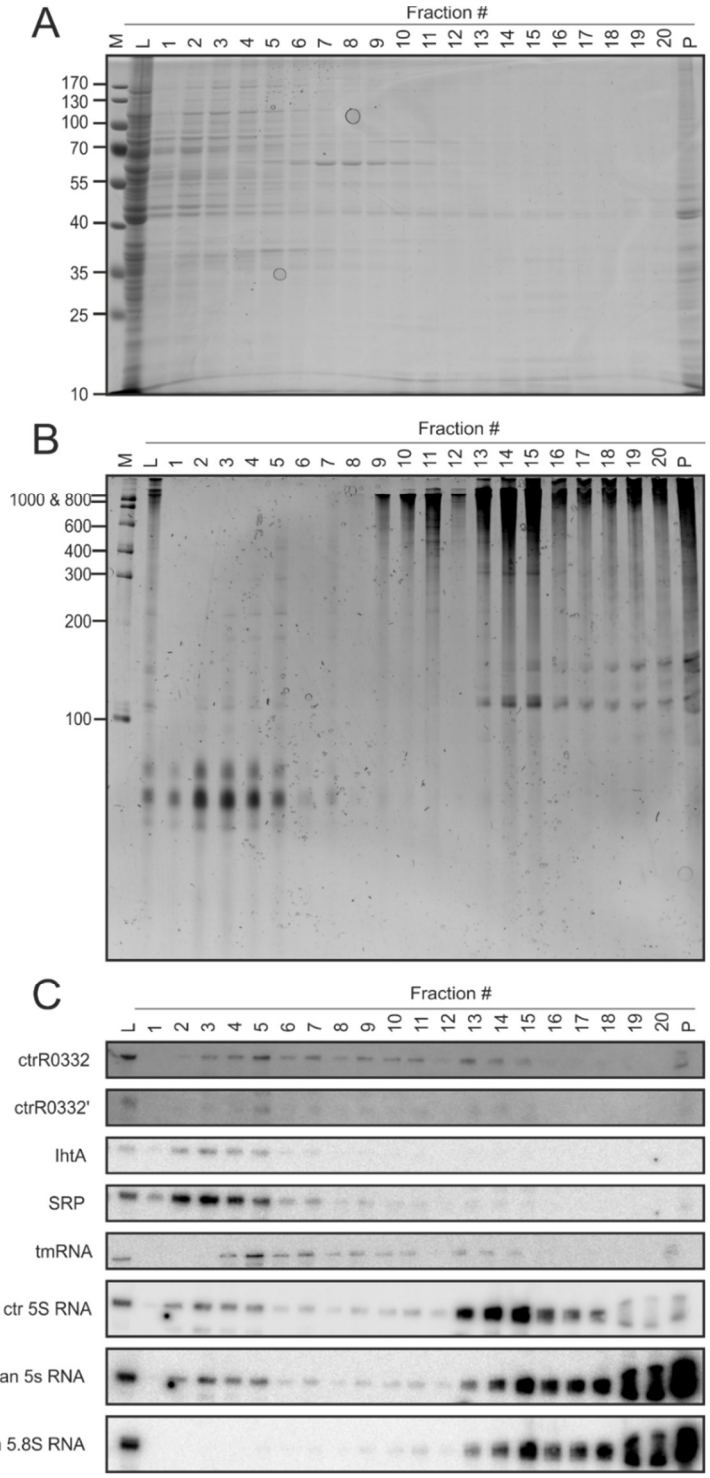
Appendix figure 41: Results from third replicate using following conditions: pellets from forty 150 cm^2-dishes of Hela229 cells infected with *Chlamydia* were lysed in a FastPrep at 4.0 m/s for 20 s once. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. (A) Proteins were separated using a 12 % SDS-PAGE and visualised via silver staining. The first lane shows the protein marker (M) followed by a lysis control and the samples from the different fractions of the gradient collected from top (1) to pellet (P). (B) RNAs were loaded onto 7M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows Riboruler Low range, followed by the lysis control and the samples of the different fractions of the gradient collected from top (1) to pellet (P). (C) RNAs were separated on a 7M Urea 6% PAGE and transferred to a nylon membrane before probing for ctrR0332. The processed form of ctrR0332 was detected at ~80 bp (ctrR0332'). The order of the samples is as described in B, but without marker.



Appendix figure 42: Absorbance of all collected fractions (1-P) measured at 260nm via NanoDrop spectrophotometer. Pellets of forty 150 cm^2 dishes of cells infected with *Chlamydia* were used and lysed at 4 m/s for 20 s once.



Appendix figure 43: Results from second replicate with following conditions: chlamydial pellets from forty 150 cm^2-dishes of infected Hela229 cells were lysed using 0.1 mm silica beads by 15 s of vortexing and 15 s cooling on ice for 5 times. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. (A) Proteins were separated using a 12 % SDS-PAGE and visualised via colloidal Coomassie staining. The first lane shows the protein marker (M) followed by a lysis control and the samples from the different fractions of the gradient collected from top (1) to pellet (P). (B) RNAs were loaded onto 7M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows RiboRuler Low range, followed by the lysis control and samples from the different fractions of the gradient collected from top (1) to pellet (P). (C) RNAs were again separated on a 7M Urea 6% PAGE but then transferred to a nylon membrane before probing for the RNAs: ctrR0332 and IhtA, signal recognition particle (SRP), transfer-messenger RNA (tmRNA), chlamydial 5S RNA, human 5S RNA and human 5.8S RNA. The processed form of ctrR0332 was detected at ~80 bp (ctrR0332'). The order of the samples is as described in B, but without marker.



Appendix figure 44: Results from third replicate with following conditions: chlamydial pellets from forty 150 cm^2-dishes of infected Hela229 cells were lysed using 0.1 mm silica beads by 15 s of vortexing and 15 s cooling on ice for 5 times. The lysate was separated on a 10%-40% glycerol gradient before samples were taken. (A) Proteins were separated using a 12 % SDS-PAGE and visualised via colloidal Coomassie staining. The first lane shows the protein marker (M) followed by a lysis control and the samples from the different fractions of the gradient collected from top (1) to pellet (P). (B) RNAs were loaded onto 7M Urea 6% PAGE and visualized via ethidium bromide. The first lane shows Riboruler Low range, followed by the lysis control and samples from the different fractions of the gradient collected from top (1) to pellet (P). (C) RNAs were again separated on a 7M Urea 6% PAGE but then transferred to a nylon membrane before probing for the RNAs: ctrR0332 and IhtA, signal recognition particle (SRP), transfer-messenger RNA (tmRNA), chlamydial 5S RNA, human 5S RNA and human 5.8S RNA. The processed form of ctrR0332 was detected at ~80 bp (ctrR0332'). The order of the samples is as described in B, but without marker.

7.4.5. Grad-seq trimming

Following command was used to trim the raw data of the gradient high-throughput sequencing:

```
cutadapt \
    -q 20 \
    -m 1 \
    -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC \
    -o ... \
```

7.4.6. Grad-seq analysis script

The following script was written to analyse the high-throughput data from the gradient:

```
import argparse
import pandas as pd
import os
from scipy import stats
from sklearn.cluster import KMeans
from sklearn.cluster import AgglomerativeClustering
from sklearn.manifold import TSNE
import textwrap
from tslearn.clustering import KShape
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt

parser = argparse.ArgumentParser(description='Calculation of correlations between one
RNA from the RNA data and the proteins')
parser.add_argument('--rnafile', '-rf',help='File with Gradatient RNAdata', required=True)
parser.add_argument('--proteinfile', '-pf',help='File with Gradatient proteindata',
required=True)
parser.add_argument('--selector', '-sf',help='obtain selected Features use features with
space: CDS tRNA rRNA sRNA hCDS hsProtein ctrProtein',nargs='+')
parser.add_argument('--table_start_rna', '-tsr',help='table start of RNAtable',type=int,
required=True)
parser.add_argument('--table_end_rna', '-ter',help='table end of RNAtable',type=int,
required=True)
parser.add_argument('--table_start_protein', '-tsp',help='table start of proteintable',type=int,
required=True)
parser.add_argument('--table_end_protein', '-tep',help='table end of proteintable',type=int,
required=True)
parser.add_argument('--principle', '-pca',help='how many Principle Components',type=int)
parser.add_argument('--output', '-o',help='basename with the correletation between the RNA
and the proteins from the Gradient')
args = parser.parse_args()
print("Following Parameters were set:")
```

```
print("RNA-File:",args.rnafile)
print("Protein-File:",args.proteinfile)
print("RNA table start:",args.table_start_rna)
print("RNA table end:",args.table_end_rna)
print("Protein table start",args.table_start_protein)
print("Protein table end:",args.table_end_protein)
if str(args.principle) != "None":
        print("Principle Components:",args.principle)

#output=str(output)[2:-3]
if str(args.output) == "None":
        args.output="Data.csv"
print("Outputfile path",args.output)
if str(args.selector) != "None":
        print("Selected Features",args.selector)

###reading RNA file frome Gradient
###creation of one Table
####read rnafile (type,identifier,annotation,values)
def _read_rnafile(filename,Valuestart,Valueend):
        rnafile=pd.read_table(filename)
        table=pd.concat([rnafile.ix[:,['Feature']],rnafile.ix[:,['Gene']],rnafile.ix[:,['Attributes']],rnafi
le.iloc[:,int(Valuestart):int(Valueend)]],axis=1)
        table.columns=
['Feature','Gene','Attributes','1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19','
20','P']
        return table
####read protein (type,identifier,annotation,values)
def _read_proteinfile(filename,Valuestart,Valueend):
        proteinfile=pd.read_table(filename)
        proteinfile['Feature']=_identify_chlamydial_proteins(proteinfile.ix[:,'Fasta.headers'])
        table=pd.concat([proteinfile.ix[:,['Feature']],proteinfile.ix[:,['Gene.names']],proteinfile.ix[
:,['Protein.names']],proteinfile.iloc[:,int(Valuestart):int(Valueend)]],axis=1)
        table.columns=
['Feature','Gene','Attributes','1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19','
20','P']
        return table
###endfile
def _identify_chlamydial_proteins(Fastaheaders):
        idlist=Fastaheaders
        id_list=[]
        for x in idlist:
                if 'OS=Chlamydia' in x:
                        id_list.append('ctrProtein')
                else:
                        id_list.append('hsProtein')
        return id_list

def _append_rna_protein_table(rnatable,proteintable):
```

```
        return rnatable.append(proteintable)

####tablefeature selector
def _select_features(table,arguments):
        selected_table=table
        return selected_table[selected_table["Feature"].isin(arguments)]

###clustering

def _extract_valuetable(Valuetable):
        return Valuetable.loc[:,'1':'P']

def _kmeans_clustering (Valuetable,number_of_clusters):
        Valuetable.as_matrix()
        k_means = KMeans(n_clusters=number_of_clusters,random_state=0)
        k_means.fit(Valuetable)
        return k_means.labels_

def _hierarchical_clustering (Valuetable,number_of_clusters):
        Valuetable.as_matrix()
        h_clust =
AgglomerativeClustering(n_clusters=number_of_clusters,affinity="euclidean",linkage="ward")
        h_clust.fit(Valuetable)
        return h_clust.labels_

def _kShape_clustering(Valuetable,number_of_clusters):
        Valuetable.as_matrix()
        kshape = KShape(n_clusters=number_of_clusters, random_state=0)
        kshape.fit(Valuetable.as_matrix())
        return kshape.labels_

def _do_both_clusterings_and_add_to_table(Valuetable,minclusters,maxclusters):
        values=_extract_valuetable(Valuetable)
        for i in range(minclusters,maxclusters,1):
                Valuetable[str(i)+'_Kmeans_cluster']=_kmeans_clustering(values,i)
        for i in range(minclusters,maxclusters,1):
                Valuetable[str(i)+'_hierarchical_cluster']=_hierarchical_clustering(values,i)
        for i in range(minclusters,maxclusters,1):
                Valuetable[str(i)+'_kShape_cluster']=_kShape_clustering(values,i)

###tsne with normalizied values
def _do_tsne(valuetable,perplexity=30):
        values=_extract_valuetable(valuetable)
        tsne = TSNE(n_components=2, random_state=0,
perplexity=perplexity).fit_transform(values)
        #np.set_printoptions(suppress=True)
        return pd.DataFrame(tsne,columns=['TSNE_1','TSNE_2'])

def _merge_value_cluster_dataframe_and_tsne(valuecluster,tsne):
```

```
        valuecluster['TSNE_1']=tsne.ix[:,['TSNE_1']]
        valuecluster['TSNE_2']=tsne.ix[:,['TSNE_2']]
        return valuecluster

###pca analyse with normalizied values

def _do_pca(valuetable,number_of_pca):
        values=_extract_valuetable(valuetable)
        pca=PCA(n_components=number_of_pca).fit_transform(values)
        columns=[]
        for i in range(1,number_of_pca+1,1):
                columns.append("pca_"+str(i))
        principalDf = pd.DataFrame(data = pca
         , columns = columns,index=values.index)
        return principalDf

## do elbows graphs

def _do_pca_elbow(valuetable):
        values=_extract_valuetable(valuetable)
        pca=PCA(n_components=20)
        pca2=pca.fit_transform(values)
        plt.plot(pca.explained_variance_ratio_.tolist())
        plt.xlabel('number of components')
        plt.ylabel('cumulative explained variance')
        plt.tight_layout()
        plt.savefig(args.output + "_pca.pdf", format='pdf')
        plt.close()

def _do_kmeans_elbow(valuetable):
        values=_extract_valuetable(valuetable)
        sse = (Smirnov et al., 2016a)
        for k in range(1, 10):
           kmeans = KMeans(n_clusters=k, max_iter=1000).fit(values)
           sse[k] = kmeans.inertia_
        plt.plot(list(sse.keys()), list(sse.values()))
        plt.xlabel("Number of cluster")
        plt.ylabel("SSE") #https://stackoverflow.com/questions/19197715/scikit-learn-k-
means-elbow-criterion
        plt.tight_layout()
        plt.savefig(args.output + "_kmeans.pdf", format='pdf')
        plt.close()

####testcasd#
testtable =
_append_rna_protein_table(_read_rnafile(args.rnafile,args.table_start_rna,args.table_end_rn
a),_read_proteinfile(args.proteinfile,args.table_start_protein,args.table_end_protein))
if str(args.selector) != "None":
        testtable=_select_features(testtable,args.selector)
```

```
_do_both_clusterings_and_add_to_table(testtable,3,10)
testtable=_merge_value_cluster_dataframe_and_tsne(testtable,_do_tsne(testtable))
if str(args.principle) != "None":
        testtable=testtable.join(_do_pca(testtable, args.principle))
testtable.to_csv(args.output,sep="\t")


_do_pca_elbow(testtable)
_do_kmeans_elbow(testtable)
```

8. Danksagung